

Tuần 4 + 5: Định danh vi sinh vật và xác định sự có mặt của gen chức năng ở vi sinh vật

Tuần 4: Xử lý trình tự gen mã hóa 16S rRNA và định danh vi sinh vật

Vi sinh vật trong môi trường có nhiều tiềm năng ứng dụng nhờ khả năng trao đổi chất đa dạng, có khả năng sinh trưởng và phát triển với tốc độ lớn, nhu cầu dinh dưỡng đơn giản, có thể sử dụng nguồn nguyên liệu, phế phụ phẩm rẻ tiền từ công nghiệp và nông nghiệp. Việc khai thác vi sinh vật trong môi trường cho các ứng dụng sản xuất enzyme, các sản phẩm có hoạt tính sinh học hoặc trong xử lý chất thải trong môi trường đều bắt đầu từ phân lập chúng từ môi trường. Sau đó các đặc điểm sinh hóa và phân loại chúng được thực hiện trước khi có thể khai thác chúng.

Trong các bài thực hành ở hai tuần cuối này, bạn sẽ định danh vi khuẩn bằng so sánh trình tự đoạn gen mã hóa 16S rRNA với các trình tự gen mã hóa 16S rRNA trong cơ sở dữ liệu NCBI, và thiết kế mồi cho phản ứng PCR phát hiện sự có mặt của một đoạn gen quy định tính trạng quan tâm của chủng vi khuẩn đó.

1. Bạn phân lập được một chủng vi khuẩn và định danh nó bằng cách khuếch đại đoạn gen mã hóa 16S rRNA (kích thước khoảng 1500 bp) và xác định trình tự của nó bằng phương pháp Sanger sequencing hai chiều. Kết quả của giải trình tự là hai trace file thể hiện trình tự được xác định chiều xuôi sử dụng mồi 27F và chiều ngược sử dụng mồi 1492R. Hai file được đặt tên **tên_chủng_27F.ab1** và **tên_chủng_1492R.ab1** có trên Teams của nhóm thí nghiệm.

a) Từ hai trình tự này, tạo trình tự consensus của đoạn gen mã hóa 16S rRNA của chủng.

Hướng dẫn cách tạo trình tự consensus bằng MEGA12 và Bioedit

- Mở phần mềm MEGA12, nhấn trái chuột vào **Align**, chọn **Edit/View Sequencer File (Trace)**, chọn file trình tự từ đầu 27F, cửa sổ **TraceEditor** hiện ra, bạn kiểm tra cẩn thận các tín hiệu đọc, mức độ nhiễu, chọn 25 – 30 nucleotide đầu tiên và chọn từ nucleotide thứ 850 trở đi và xóa các đoạn được chọn bằng cách ấn nút Delete trên bàn phím. Trên cửa sổ **TraceEditor**, chọn **File/Export FASTA** file để lưu file fasta của trình tự từ đầu 27F
- Mở trình tự từ đầu 1492R tương tự như bước trên, sau khi đã xóa các vùng trình tự nhiễu, chọn **Edit/Reverse Complement**, sau đó lưu file fasta tương tự như bước trên.
- Mở **Bioedit**, chọn **File/New Alignment**, cửa sổ alignment hiện ra. Sau đó chọn **File/Import/Sequence alignment file**, chọn 2 trình tự fasta được tạo ra ở hai bước trên.
- So sánh hai trình tự với nhau theo các bước sau: Chọn **Edit>Select all sequences** (hoặc Ctrl+A), khi đó hai trình tự sẽ được highlighted. Chọn **Sequence/Pairwise alignment/Align two sequences (allow ends to slide)**, sau khi thực hiện

alignment, có hai cửa sổ mới xuất hiện: cửa sổ output có các trình tự được align với nhau và cửa sổ tổng kết các thông số của output.

- Tạo trình tự consensus: ấn chuột vào cửa sổ output ở bước trên, sau đó chọn **Accessory Application/CAP contig assembly program**, cửa sổ CAP contig assembly program interface mở ra, ấn **Run Application**, sau đó Windows Terminal hiện ra thông báo các trạng thái của chương trình CAP contig assembly (initiate, make, pair, assemble, repair, form tree và show kết quả, con chuột nháy ở dòng cuối của Terminal, ấn **enter**, cửa sổ output của assembly xuất hiện, trong đó trình tự consensus là trình tự thứ ba).
- Lưu trình tự consensus gen mã hóa 16S rRNA: ấn chuột vào cửa sổ output của contig assembly, chọn trình tự consensus này, sau đó chọn **Edit/copy sequences to clipboard (Fasta format)**, mở app Notepad, paste trình tự và lưu file.

Copy trình tự dạng fasta xuống ô trống dưới đây

>Contig-0

```
GAACGCACCGCAAGGTGAGTGGCAGACGGGTGAGTAACACGTGGGATCTGCCAATGGTACGGAATAGCTCCG  
GGAAACTGGAATTAAATACCGTATGAGCCCGCAAGGGAAAGATTATGCCATTGGATGAACCGCGTCGGATT  
AGCTAGTTGGTGTGGTAAAGGCGCACCAAGGCGACGATCCGTAGCTGGTCTGAGAGGATGATCAGCCACACTGG  
GACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGAAATATTGGACAATGGCGCAAGCCTGATCCA  
GCCATGCCCGTGAGTGTAGAAGGCCTTAGGGTTGAAAGCTTTGCCAGAAGATAATGACGGTAGTCGG  
AGAAGAAGCCCCGGCTAACCTCGTGCAGCAGCCGGTAATACGAAGGGGGTAGCGTTGGAATCAGTGG  
GGCGTAAAGCGCACGTAGGCGGATTGTTAAGTCAGGGGTGAAAGCCTGGAGCTCAACTCCAGAACTGCCCTGA  
TACTGGCAATCTCGAGTCCGGAAAGAGGTAAGTGGAACTGCGAGTGTAGAGGTAAGCTGAGATATTGCAAG  
AACACCAGTGGCGAAGGCGGTTACTGGTCCGGTACTGACGCTGAGGTGCGAAAGCGTGGGAGCAAACAGGAT  
TAGATACCTGGTAGTCCACGCCGTAAACGATGGAGGCTAGCCGGTGGCGAGCATGCTCGTCACTGGCGCAGCT  
AACGCATTAAGCCTCCGCCTGGGAGTACGGTCGAAGATTAACGAACTCAAAGGAATTGACGGGGCCCGACA  
AGCGGTGGAGCATGTGGTTAACCGCAAGCAACGCCAGAACCTTACCGCCTTGACATGCCCGGAATTGGA  
TCAGAGATGAACCAAGCTCTCGGAGGCCGGGACACAGGTGCTGCATGGCTGTCGTCACTCGTGTGAGAT  
GTTGGGTTAACGAGCTCCCGCAACGAGCGAACCCCTGCCCTAGTTGCCATCATTCAAGTGGCAGCTAGGGGAC  
TGCCGGTGATAAGCCGAGAGGAAGGTGGGATGACGTCAAGTCCATGCCCTACGGGCTGGCTACACACG  
TGCTACAATGGCGGTGACAGTGGACGCGAACCGCGAGGGTGAGCAAATCTCAAAGCCGCTCAGTTCGGA  
TTGCACTCTGCAACTCGAGTGCATGAAGTTGGAATCGTAGTAATCGTGGATCAGCACGCCACGGTGAATACGT  
TCCCGGGCCTGTACACACCGCCGTACACCATGGAGTTGGCTTACCGAAGGCCTGCCTAACCGCAA
```

2. Sử dụng công cụ blastn của NCBI để xác định danh tính của chủng vi khuẩn này, với các thông tin sau:

- + Cơ sở dữ liệu để so sánh: rRNA/ITS databases/16S ribosomal RNA sequences (bacteria and archaea)
- + Exclude: chọn Uncultured/environmental sample sequences

a) Lập bảng thống kê sự tương đồng về trình tự và kết luận về danh tính của chủng với 5 trình tự đầu tiên khi sắp xếp kết quả tăng dần giá trị E-value

Tên loài/chủng	Tỷ lệ tương đồng (%)	Mã hiệu trình tự
<i>Ancylobacter novellus</i> DSM 506	99.25	NR_074219.1
<i>Ancylobacter koreensis</i>	99.10	NR_113962.1
<i>Ancylobacter koreensis</i>	99.10	NR_041013.1
<i>Ancylobacter novellus</i> DSM 506	98.50	NR_025859.1
<i>Ancylobacter lacus</i>	97.97	NR_180559.1

Kết luận:

b) Sử dụng cơ sở dữ liệu Taxonomy trên NCBI, bạn hãy cho biết vị trí phân loại của chủng nghiên cứu

Hypomicrobiales / Xanthobacteraceae / Ancylobacter

3. Xây dựng cây phát sinh loài thể hiện mối quan hệ tiến hóa của chủng với các chủng có quan hệ gần gũi.

Thu thập các trình tự mã hóa 16S rRNA (ít nhất 10 trình tự) của các chủng có quan hệ loài gần gũi (trong cùng một họ - **family**) và của một chủng không thuộc ngành của chủng nghiên cứu để làm **outgroup** cho cây phát sinh loài.

a) Tải trình tự fasta và lưu vào các file fasta riêng biệt hoặc lưu vào cùng một file.

Copy các trình tự dạng fasta vào ô trống dưới đây

(*các trình tự này là bản đã được alignment bằng MEGA):

```
#mega
!Title all_seq_final;
!Format DataType=DNA indel=- CodeTable=Standard;

!Domain=Data property=Coding CodonStart=1;
#Contig-0
GAACGCA--CCGCAAG-----GTGAGTGGCAGACGGGTGAGTAACACGTGGG
GATCTGCCAATGGTACGGAATAGCTCCGGAAACTGGAATTAAATACCGTATGAGCCGC
AAGGGGAAAGA-----TTTATGCCATTGGATGAACCCGCGTCGGATTAGC
```

TAGTTGGTGTGG-TAAAGGCGCACCAAGGCACGATCCTAGCTGGTCTGAGAGGATGAT
CAGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGAATAT
TGGACAATGGCGCAAGCCTGATCCAGCCATGCCCGTGAGTGATGAAGGCCTTAGGTT
GTAAAGCTTTCGCCGACGAAGA-----TAATGACGGTA
GTCGGAGAAGAACGCCCCGGCTAACCTCGTGCCAGCAGCCGGTAATACGAAGGGGCTA
GCGTTGTCGGAATCACTGGCGTAAAGCG--CACGTAGGCGGATTGTTAAGTCAGGGT
GAAAGCCTGGAGCTCAACTCCAGAACTGCCCTGATACTGGCAATCTGAGTCCCGAAGA
GGTAAGTGGAACTGCGAGTGTAGAGGTGAAATTCTGTAGATATTGCAAGAACACCAGTGG
CGAAGGCGGCTTACTGGTCCGGTACTGACGCTGAGGTGCGAAAGCGTGGGAGCAAACAG
GATTAGATAACCTGGTAGTCCACGCCGTAAACGATGGAGGCTAGCC-GTTGGCGAGCATG
CTCGTCAGTGGCGCAGCTAACGCAATTAGCCTCCGCCTGGGAGTACGGTCGCAAGATT
AAAACCTCAAAGGAATTGACGGGGGCCGACAAGCGGTGGAGCATGTGGTTAATTGAA
GCAACCGCGAGAACCTTACCGCCCTTGACATGCCCGAATTGGATCAGAGATGAACCA
AGCTCTT-CGG-AGCGGGGACACAGGTGCTGCATGGCT-GTCGTCAGCTCGTGTG
GATGTTGGGTTAAGTCCCGAACGAGCGAACCCCTGCCCTAGTTGCCATCATTCA
GGGCACTCTAGGGGACTGCCGGTATAAGCCAGAGGAAGGTGGGATGACGTCAAGTC
CTCATGCCCTTACGGCTGGCTACACACGTGCTACAATGGCGGTGACAGTGGACCG
AACCCCGAGGGTGAGCAAATCTCCAAAAGC-CGTCTCAGTTGGATTGCACTCTGCAAC
TCGAGTGCATGAAGTTGAAATCGCTAGTAATCGTGGATCAGCACGCCACGGTAAACGT
TCCCGGCCTGTACACACCGCCGTACACCATGGAGTTGGCTTACCGAAGGCCT
GCGCTAACCC-GCAA

#NR_074219.1_Ancylobacter_novellus_DSM_506_16S_ribosomal_RNA_partial_sequence
GAACGCA--CCGCAAG-----GTGAGTGGCAGACGGGTGAGTAACACGTGG
GATCTGCCCAATGGTACGGAATAGCTCCGGAAACTGGAATTAAATACCGTATGTGCC
AAGGGGAAAGA-----TTTATGCCATTGGATGAACCCGCGTCGGATTAGC
TAGTTGGTGTGG-TAAAGGCGCACCAAGGCACGATCCGTAGCTGGTCTGAGAGGATGAT
CAGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGAATAT
TGGACAATGGCGCAAGCCTGATCCAGCCATGCCCGTGAGTGATGAAGGCCTTAGGTT
GTAAAGCTTTGCCGACGAAGA-----TAATGACGGTA
GTCGGAGAAGAACGCCCCGGCTAACCTCGTGCCAGCAGCCGGTAATACGAAGGGGCTA
GCGTTGTTCGGAATCACTGGCGTAAAGCG--CACGTAGGCGGATTGTTAAGTCAGGGT
GAAAGCCTGGAGCTCAACTCCAGAACTGCCCTGATACTGGCAATCTGAGTCCCGAAGA

GGTAAGTGGAACTGCGAGTGTAGAGGTGAAATTCTGTAGATATTGCAAGAACACCAGTGG
CGAAGGCGGCTTACTGGTCCGGTACTGACGCTGAGGTGCGAAAGCGTGGGAGCAAACAG
GATTAGATACCTGGTAGTCACGCCGTAAACGATGGAGGCTAGCC-GTTGGTGAGCATG
CTCATCAGTGGCGCAGCTAACGCATTAAGCCTCCGCCTGGGAGTACGGTCGCAAGATT
AAAACCTCAAAGGAATTGACGGGGCCCGACAAGCGGTGGAGCATGTGGTTAACCGAA
GCAACGCGCAGAACCTTACCGCCTTGACATGTCTCGGAATTGGACCAGAGATGGACCA
AGCTCTT-CGG-AGCCGGAACACAGGTGCTGCATGGCT-GTCGTCAGCTCGTGTGTA
GATGTTGGGTTAAGTCCCACAGAGCGAACCCCTGCCCTAGTTGCCATTAAGTT
GGGCACTCTAGGGGACTGCCGGTATAAGCCGAGAGGAAGGTGGGATGACGTCAAGTC
CTCATGGCCCTTACGGGCTGGCTACACACGTGCTACAATGGCGGTGACAGTGGGATGCG
AACCCCGCAGGGTGAGCAAATCTCCAAAAGC-CGTCTCAGTCGGATTGCACTCTGCAAC
TCGAGTGCATGAAGTTGAAATCGCTAGTAATCGTGGATCAGCACGCCACGGTGAATACGT
TCCC GG CTT GTACACACC CGCC GT CA ACC ATGGAGTTGGCTTACCGAAGGGCCT
GCGCTAACCC-GCAA

#NR_113962.1_Ancylobacter_koreensis_strain_NBRC_100963_16S_ribosomal_RNA_partial_sequence
GAACGCA--CCGCAAG-----GTGAGTGGCAGACGGGTGAGTAACACGTGGG
GATCTGCCCAATGGTACGGAATAGCTCCGGAAACTGGAATTAAATACCGTATGTGCCCTT
CGGGGGAAAGA-----TTTATGCCATTGGATGAACCCGCGTCGGATTAGC
TAGTTGGTGTGG-TAAAGGCGCACCAAGGCGACGATCCGTAGCTGGCTGAGAGGATGAT
CAGCCACACTGGGACTGAGACACGGCCAGACTCCTACGGGAGGCAGCAGTGGGAATAT
TGGACAATGGCGCAAGCCTGATCCAGCCATGCCCGTGAGTGATGAAGGCCTTAGGGTT
GTAAAGCTTTGCCGACGAAGA-----TAATGACGGTA
GTCGGAGAAGAAGCCCCGGCTAACCTCGTGCAGCAGCGCGGTAAATACGAAGGGGCTA
GCGTTGTTCGGAATCACTGGCGTAAAGCG--CACGTAGGCGGATTGTTAAGTCAGGGT
GAAAGCCTGGAGCTCAACTCCAGAACTGCCCTGATACTGGCAATCTGAGTCCGGAAGA
GGTAAGTGGAACTGCGAGTGTAGAGGTGAAATTCTGTAGATATTGCAAGAACACCAGTGG
CGAAGGCGGCTTACTGGTCCGGTACTGACGCTGAGGTGCGAAAGCGTGGGAGCAAACAG
GATTAGATACCTGGTAGTCACGCCGTAAACGATGGAGGCTAGCC-GTTGGTGAGCATG
CTCATCAGTGGCGCAGCTAACGCATTAAGCCTCCGCCTGGGAGTACGGTCGCAAGATT
AAAACCTCAAAGGAATTGACGGGGCCCGACAAGCGGTGGAGCATGTGGTTAACCGAA
GCAACGCGCAGAACCTTACCGCCTTGACATGTCCGGAATTGGATCAGAGATGGACCA

AGCTCTT-CGG-AGCCGGGAACACAGGTGCTGCATGGCT-GTCGTCAGCTCGTGTGTA
GATGTTGGGTTAAGTCCCACGAGCGAACCGCTGCCCTAGTTGCCATTCAGTT
GGCACTCTAGGGGACTGCCGGTATAAGCGAGAGGAAGGTGGGATGACGTCAAGTC
CTCATGGCCCTTACGGCTGGCTACACACGTGCTACAATGGCGGTGACAGTGGAAAGCG
AACCGCGAGGGTGAGCAAATCTCCAAAAGC-CGTCTCAGTCGGATTGCACTCTGCAAC
TCGAGTGCATGAAGTTGAAATCGCTAGTAATCGTGGATCAGCACGCCACGGTGAATACGT
TCCCAGGCTTGTACACACCGCCCGTACACCATGGAGTTGGTTTACCGAAGGCCT
GCGCTAACCC-GCAA

#NR_180560.1_Ancylobacter_plantiphilus_strain_1TC_16S_ribosomal_RNA_partial_sequence
GAACGCC--CCGCAAG-----GGGAGTGGCAGACGGGTGAGTAACACGTGGG
GATCTGCCAATGGTACGGAATAGCTCCGGAAACTGGAATAATACCGTATAAGCCGC
AAGGGGAAAGA-----TTTATGCCATTGGATGAACCCGCGTCGGATTAGC
TAGTTGGTGGGG-TAAAGGCCTACCAAGGCAGCGATCCGTAGCTGGTCTGAGAGGATGAT
CAGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGAGGCAGCAGTGGGAATAT
TGGACAATGGCGCAAGCCTGATCCAGCCATGCCCGTGAGTGATGAAGGCCTTAGGTT
GTAAAGCTTTGCCGACGAAGA-----TAATGACGGTA
GTCGGAGAAGAAGCCCCGGCTAACCTCGTGCCAGCAGCCGGTAATACGAAGGGGCTA
GCGTTGTCGGAATCACTGGCGTAAAGCG--CACGTAGGGATTGTTAAGTCAGAGGT
GAAAGCCTGGAGCTCAACTCCAGAACTGCCCTGATACTGGAATCTCAGTCCGGAAAGA
GGTAAGTGGAACTGCGAGTGTAGAGGTGAAATTCTGTAGATATTGCAAGAACACCAGTGG
CGAAGGCGCTTACTGGTCCGGTACTGACGCTGAGGTGCGAAAGCGTGGGAGCAAACAG
GATTAGATAACCTGGTAGTCCACGCCGTAAACGATGGAGGCTAGCC-GTTGGTGAGCATG
CTCATCAGTGGCGCAGCTAACGCATTAAGCCTCCGCCCTGGGAGTACGGTCGCAAGATT
AAAACCTAAAGGAATTGACGGGGCCCGACAAGCGGTGGAGCATGTGGTTAATTGAA
GCAACCGCGAGAACCTTACCGCCTTGACATGCCAGGACAGTTACAGAGATGGTT
CTCTCTT-CGG-AGCCTGGACACAGGTGCTGCATGGCT-GTCGTCAGCTCGTGTGTA
GATGTTGGGTTAAGTCCCACGAGCGAACCGCTGCCCTAGTTGCCATTCAGTT
GGCACTCTAGGGGACTGCCGGTATAAGCGAGAGGAAGGTGGGATGACGTCAAGTC
CTCATGGCCCTTACGGCTGGCTACACACGTGCTACAATGGCGGTGACAATGGGATGCA
AGACCGCGAGGTTGAGCAAATCTCCAAAAGC-CGTCTCAGTCGGATTGCACTCTGCAAC
TCGAGTGCATGAAGTTGAAATCGCTAGTAATCGTGGATCAGCACGCCACGGTGAATACGT
TCCCAGGCTTGTACACACCGCCCGTACACCATGGAGTTGGTTTACCGAAGGCCTT

GCGCTAACCC-GCAA

#NR_042794.1_Ancylobacter_vacuolatus_strain_DSM_1277_16S_ribosomal_RNA_partial_sequence

GAACGCC--CCGCAAG-----GGGAGTGGCAGACGGGTGAGTAACACGTGGG
GATCTGCCAATGGTACGGAATAATTGGAAACTGGGACTAATACCGTATGAGCCGC
AAGGGGAAAGA-----TTTATGCCATTGGATGAACCCGCGTCGGATTAGC
TAGTTGGTGTGG-TAAAGGCACCAAGGCACGGATCCGTAGCTGGTCTGAGAGGATGAT
CAGCCACACTGGGACTGAGACACGGCCA?ACTCCTACGGAGGCAGCAGTGGGAATAT
TGGACAATGGCGCAAGCCTGATCCAGCCATGCCCGTGAGTGATGAAGGCCTTAGGTT
GTAAAGCTTTGCCGACGAAGA-----TAATGACGGTA
GTCGGAGAAAAGCCCCGGCTAACCTCGTGCACGCCGGTAATACGAAGGGGCTA
GCGTTGTCGGAATCACTGGCGTAAAGCG--CACGTAGGCGGATATTAAAGTCAGGGT
GAAAGCCTGGAGCTCAACTCCAGAACTGCCCTGATACTGGTATCTGAGTCCGGAAGA
GGTAAGTGGAACTGCGAGTGTAGAGGTGAAATTCTGTAGATATTGCAAGAACACCAGTGG
CGAAGGCGGCTTACTGGTCCGGTACTGACGCTGAGGTGCGAAAGCGTGGGAGCAAACAG
GATTAGATACCCCTGGTAGTCCACGCCGTAAACGATGGAGGCTAGCC-GTTGGTGAGCATG
CTCATCAGTGGCGCAGCTAACGCATTAAGCCTCCGCCTGGGAGTACGGTCGCAAGATT
AAAACCTCAAAGGAATTGACGGGGCCCGACAAGCGGTGGAGCATGTGGTTAATTGAA
GCAACCGCGAGAACCTTACCGCCTTGACATGTCGGACGGTTACAGAGATGGTTTC
TTCTCTT-CGG-AGCCGGAACACAGGTGCTGCATGGCT-GTCGTAGCTCGTGTGTA
GATGTTGGGTTAAGTCCCACGAGCGAACCCCTGCCCTAGTTGCCATCATTAAGTT
GGGCACTCTAGGGGACTGCCGGTATAAGCCGAGAGGAAGGTGGGATGACGTCAAGTC
CTCATGCCCTTACGGCTGGCTACACACGTGCTACAATGCCGGTACAGTGGGAAGCG
AACCTGCGAGGGTAAGCAAATCTCCAAAGC-CGTCTCAGTCGGATTGCACTCTGCAAC
TCGAGTGCATGAAGTTGGAATCGCTAGTAATCGGGATCAGCATGCCACGGTGAATACGT
TCCCGGGCCTGTACACACCGCCGTACACCATGGAGTTGGTTTACCGAAGGCCT
GCGCTAACCC-GCAA

#NR_044737.2_Ancylobacter_aquaticus_strain_ATCC_25396_16S_ribosomal_RNA_partial_sequence

GAACGCC--CCGCAAG-----GGGGGTGGCAGACGGGTGAGTAACACGTGGG
GATCTGCC?AATGGTACGGAATAATTGGAAACTGGGACT?ATACCGTATGAGCCGC
AAGGGGAAAGA-----TTTATGCCATTGGATGAACCCGCGTCGGATTAGC

TAGTTGGTGTGG-TAAAGGCGCACCAAGGCAGCATCCGTAGCT?GTCTGAGAGGATGAT
CAGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGAATAT
TGGACAATGGCGCAAGCC?ATCCAGCCATGCCCGTGAGTGATGAAGGCCTTAGGGTT
GTAAAGCTTTCGCCGACGAAGA-----TAATGACGGTA
GTCGGAGAAGAAGCCCCGGCTAACCTCGTGCCAGCAGCCGGTAATACGAAGGGGC?A
GCGTTGTCGGAATCACTGGCGTAAAGCG--CACGTAGGCGGATATTAAAGTCAGGGT
GAAAGCCTGGAGCTCAACTCCAGAACTGCCCTGATACTGGGTATCTGAGTCCCGAAGA
GGTAAGTGBACTGCGAGTGTAGAGGTGAAATTCTGTAGATATTGCAAGAACACCAGTGG
CGAAGGCGGCTTACTGGTCCGGTACTGACGCTGAGGTGC?AAAGCGTGGGAGCAAACAG
GATTAGATAACCTGGTAGTCCACGCCGTAAACGATGGAGGCTAGCC-GTTGGTGAGCATG
CTCATCAGTGGCGCAGCTAACGATTAAGCCTCCCGCTGGGAGTACGGTCGCAAGATT
AAAACCTCAAAGGAATTGACGGGGCC?GCACAAGCGGTGGAGCATGTGGTTAATTGAA
??AACGCGCAGAACCTTACCGCCCTTGACATGTCCCGACGGTTACAGAGATGGTTTC
TTCTCTT-CGG-AGCGGGAACACAGGTGCTGCATGGCT-GTCGTCAGCTCGTGTGTA
GATGTTGGGTTAAGTCCCGAACGAGCGAACCCCTGCCCTAGTTGCCATCATTAAGTT
GGGCACTCTAGGGGACTGCCGGTATAAGCCAGAGGAAGGTGGGATGACGTCAAGTC
CTCATGCCCTTACGGCTGGCTACACACGTGCTACAATGGCGGTGACAGTGGGAAGCG
AACCTGCGAGGGTAAGCAAATCTCCAAAAGC-CGTCTCAGTTGGATTGCACTCTGCAAC
TCGAGTGCATGAAGTTGAAATCGCTAGTAATCGTGGATCAGCATGCCACGGTAATACGT
TCCCGGGC?TTGTACACACCGCCCGTACACCATGGAGTT?GTTTACCCGAAGGCCT
GCGCTAACCC-GCAA

#NR_025172.1_Xanthobacter_aminoxidans_strain_14a_16S_ribosomal_RNA_partial_sequence
GAGCGCC--CAGCAAT-----GGGAGCGGCAGACGGGTGAGTAACACGTGG
GATCTACCCAATGGTACGGAATAACCCAGGGAAACTGGATTAAATACCGTATGTGCCCTT
CGGGGGAAAGA-----TTTATGCCATTGGATGAACCCGCGTCGGATTAGC
TAGTTGGTGAGG-TAAAGGCTCACCAAGGCAGCATCCGTAGCTGGTCTGAGAGGATGAT
CAGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGAATAT
TGGACAATCGCGCAAGCCTGATCCAGCCATGCCCGTGAGTGATGAAGGCCTTAGGGTT
GTAAAGCTTTGCCGGTGAAGA-----TAATGACGGTA
ACCGGAGAAGAAGCCCCGGCTAACCTCGTGCCAGCAGCCGGTAATACGAAGGGGCTA
GCGTTGCTCGGAATCACTGGCGTAAAGCG--CACGTAGGCGGATGTTAAGTCAGGGT
GAAATCCTGGAGCTCAACTCCAGAACTGCCCTGATACTGGCGATCTGAGTTGAGAGA

GGTTGGTGGAACTCCGAGTGTAGAGGTGAAATTCTGATATTGGAAAGAACACCAGTGG
CGAAGGCAGGCCAACTGGCTCGATACTGACGCTGAGGTGCGAAAGCGTGGGAGCAAACAG
GATTAGATACCTGGTAGTCACGCCGTAAACGATGGATGCTAGCCTGTGGGGGGTTA
CCTCTCAGTGGCGCAGCTAACGTCTTAAGCATCCGCCTGGGAGTACGGTCGCAAGATT
AAAACCTCAAAGGAATTGACGGGGCCCGACAAGCGGTGGAGCATGTGGTTAATTGAA
GCAACCGCGAGAACCTTACCGCCTTGACATGGCAGG-ACGACTTCCGGAGACGGATTT
CTTCCAG-CAATGGACCTCGACACAGGTGCTGCATGGCT-GTCGTCAGCTCGTGTGTA
GATGTTGGGTTAAGTCCCACAGCGCAACCGAGCGAACCTCGCCTCTAGTTGCCATTCAGTT
GGGCACTCTAGAGGGACTGCCGGTGATAAGCCGAGAGGAAGGTGGGATGACGTCAAGTT
CTCATGGCCCTTACGGGCTGGCTACACACGTGCTACAATGGCGGTGACAGTGGATGCG
AACCCCGAGGGTAAGCAAATCTCCAAAAGC-CGTCTCAGTCGGATTGCACTCTGCAAC
TCGAGTGCATGAAGTTGAAATCGCTAGTAATCGTGGATCAGCATGCCACGGTGAATACGT
TCCCCGGCCTGTACACACCGCCCGTACACCATGGAGTTGGCTTACCGAAGGGCCT
GCGCTAACCC-GCAA

#NR_179776.1_Aquabacter_cavernae_strain_Sn-9_2_16S_ribosomal_RNA_partial_sequence
GAGCGCC--CAGCAAT-----GGGAGCGGCAGACGGGTGAGTAACCGTGGG
GATCTGCCGATGGTACGGAATAATTCCGGAAACTGGGACTAATACGTATGTGCCCGC
AAGGGGAAAGA-----TTTATGCCATCGGATGAACCGCGTCGGATTAGC
TAGTTGGTGTGG-TAAAGGCGACCAAGGCACGATCCGTAGCTGGCTGAGAGGATGAT
CAGCCACACTGGACTGAGACACGGCCAGACTCCTACGGGAGGCAGCAGTGGGAATAT
TGGACAATGGCGCAAGCCTGATCCAGCCATGCCCGTGAGTGATGAAGGCCTAGGGTT
GTAAAGCTTTGCCGGTGAAGA-----TAATGACGGTA
ACCGGAGAAGAAGCCCCGGCTAACTTCGTGCCAGCAGCGCGGTAAACGAAGGGGCTA
GCGTTGCTCGGAATCACTGGCGTAAAGCG--CACGTAGGCGGTGTTAAGTCAGAGGT
GAAAGCCTGGAGCTCAACTCCAGAACTGCCTTGATACTGGCGACCTAGAGTTGAGAGA
GGTTGGTGGAACTCCGAGTGTAGAGGTGAAATTCTGATATTGGAAAGAACACCAGTGG
CGAAGGCAGGCCAACTGGCTCGATACTGACGCTGAGGTGCGAAAGCGTGGGAGCAAACAG
GATTAGATACCTGGTAGTCACGCCGTAAACGATGGATGCTAGCC-GTTGGGGAGCTTG
CTCTTCAGTGGCGCAGCTAACGCCTTAAGCATCCGCCTGGGAGTACGGTCGCAAGATT
AAAACCTCAAAGGAATTGACGGGGCCCGACAAGCGGTGGAGCATGTGGTTAATTGAA
GCAACCGCGAGAACCTTACCGCCTTGACATGGCAGG-ACGATTCCAGAGATGGATCT
CTTCCAG-CAATGGACCTGCACACAGGTGCTGCATGGCT-GTCGTCAGCTCGTGTGTA

GATGTTGGGTTAAGTCCCGAACGAGCGAACCCCTGCCTTAGTTGCCATCATTAGTT
GGGCACTCTAAAGGGACTGCCGGTGATAAGCCGCGAGGAAGGTGGGATGACGTCAAGTC
CTCATGGCCCTTACGGCTGGCTACACACGTGCTACAATGGTGGTGACAGTGGATGCG
AAAGGGCGACCTCTAGCAAATCTCCAAAAGC-CATCTCAGTCGGATTGCACTCTGCAAC
TCGAGTGCATGAAGTTGGAATCGCTAGTAATCGTGGATCAGCATGCCACGGTGAATACGT
TCCCAGGCCTTGTACACACCGCCCGTCACACCATGGGAGTTGGTTTACCCGAAGGGCCT
GCGCTAACCG--GCAA

#NR_041839.1_Azorhizobium_doebereinerae_strain_BR5401_16S_ribosomal_RNA_partial_sequence

GAGCGCC--CAGCAAT-----GGGAGCGGCAGACGGGTGAGTAACCGTGGG
GATCTGCCAATGGTACGGAATAACCCAGGGAACTTGGATTAATACCGTATGTGCCCTT
CGGGGGAAAGA-----TTTATGCCATTGGATGAACCCGCGTCGGATTAGC
TAGTTGGTGTGG-TAAAGGCGACCAAGGCACGATCCGTAGCTGGTCTGAGAGGATGAT
CAGCCACACTGGGACTGAGACACGGCCAGACTCCTACGGGAGGCAGCAGTGGGAATAT
TGGACAATGGCGCAAGCCTGATTCCAGCCATGCCCGTGTGTGATGAAGGCCTAGGGTT
GTAAAGCACTTCGCCGGTGAAGA-----TAATGACGGTA
ACCGGAGAAGAAGCCCCGGCTAACCTCGTGCAGCAGCGCGTAATACGAAGGGGCAA
GCGTTGCTCGAATCACTGGCGTAAAGCG--TACGTAGGCGGTGTTAAGTCAGGGT
GAAATCCTGGAGCTCAACTCCAGAACTGCCCTTGATACTGGCGATCTGAGTCGAGAGA
GGTTGGTGGAACTGCGAGTGTAGAGGTGAAATTCTGTAGATATTGCAAGAACACCAGTGG
CGAAGGCAGCCAAGTGGCTGATACTGACGCTGAGGTACGAAAGCGTGGGAGCAAACAG
GATTAGATAACCTGGTAGTCCACGCCGTAAACGATGGATGCTAGCC-GTAGGCAGCTTG
CTGCTTAGTGGCGCAGCTAACGCATTAAGCATCCCGCTGGGAGTACGGTCGCAAGATT
AAAACCTAAAGGAATTGACGGGGCCCGACAAGCGGTGGAGCATGTGGTTAACGAA
GCAACCGCGAGAACCTTACCGCCTTGACATGGCAGG-ACGACTTCCGGAGACGGATT
CTTCCAG-CAATGGACCTGCACACAGGTGCTGCATGGCT-GTCGTCAGCTCGTGTG
GATGTTGGGTTAAGTCCCGAACGAGCGAACCCCTGCCTTAGTTGCCATCATTAGTT
GGGCACTCTAAAGGGACTGCCGGTGATAAGCCGCGAGGAAGGTGGGATGACGTCAAGTC
CTCATGGCCCTTACGGCTGGCTACACACGTGCTACAATGGCGGTGACAGTGGATGCG
AGCCTGCGAGGGTGAGCAAATCTCCAAAAGC-CGTCTCAGTCGGATTGCACTCTGCAAC
TCGAGTGCATGAAGTTGGAATCGCTAGTAATCGTGGATCAGCATGCCACGGTGAATACGT
TCCCAGGCCTTGTACACACCGCCCGTCACACCATGGGAGTTGGTTTACCCGAAGGGCCTT

GCGCTAACCTAGCAA

#NR_115492.1_Azorhizobium_caulinodans_ORS_571_16S_ribosomal_RNA_partial_sequence
GAGCGCC--CAGCAAT-----GGGAGCGGCAGACGGGTGAGTAACCGTGGGG
GATGTGCCAATGGTGCAGATAACCCAGGGAACTTGGATTAAATACCGCATGTGCCCTT
CGGGGGAAAGA-----TTTATGCCATTGGATCAACCCGCGTCTGATTAGC
TAGTTGGTGAGG-TAAAGGCTCACCAAGGCAGCATCAGTAGCTGGTCTGAGAGGATGAT
CAGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGAATAT
TGGACAATGGCGCAAGCCTGATCCAGCCATGCCGCGTGAATGATGAAGGCCTTAGGTT
GTAAAGCTTTGCCGGTGAAGA-----TAATGACGGTA
ACCGGAGAAGAAGCCCCGGCTAACCTCGTGCAGCAGCCGGTAATACGA?GGGGCAA
GC?TTGCTCGGAATCACTGGCGTAAAGCG--CACGTAGGCGATGTTAAGTCAGGGT
GAAAGCCTGGAGCTCAACTCCAGAACTGCCCTGATACTGGCGATCTGAGTTGAGAGA
GGTTGGTGGAACTCCGAGT?TAGAGGTGAAATTCTGTAGATATTGGAAAGAACACCAGT?G
CGAAGGCGGCCAACTGGCTCGATACTGACGCTGAGGTGCGAAA??GTGGGGAGCAAACAG
GATTAGATAACCTGGTAGTCCACGCCGTAAACGATGGATGCTAGCC-GTTGGGGAGCTTG
CTCTTCAGTGGCGCA??TAACGCCTTAAGCATCCCGCTGGGAGTACGGTCGCAAGATT
AAAAACTCAAAGGAATTGACGGGGGCCGCACAAGCGGTGGAGCATGTGGTTAATTGAA
GCAACCGCGAGAACCTTACCGCCTTGACATGGCAGG-ACGACTTCCGGAGACGGATT
CTTCCAG-CAATGGACCTGCACACAGGTGCTGCATGGCT-GTCGTAGCTCGTGTGA
GATGTTGGGTTAAGTCCCACGAGCGAACCGCTTGACATGGCAGG-ACGACTTCCGGAGACGGATT
GGGCACTCTAAAGGGACTGCCGGTGATAAGCCCGAGGAAGGTGGGATGACGTCAAGTC
CTCATGGCCCTTACGGCTGGCTACACACGTGCTACAATGGCGGTGACAATGGGATGCG
AGCCT?CGAGGGTGAGCAAATCTCCAAAAGC-CGTCTCAGTCGGATTGCACTCTGCAAC
TCGAGT?CATGAAGTTGGAATCGCTAGTAATCGGATCAGCATGCCACGGTGAATACGT
TCCCGGGCCTGTACACACCG?CCGTACACCATGGAGTTGGCTTACCGAAGGCCTT
GCGCTAAC--GCAA

#NR_116004.1_Labrys_wisconsinensis_strain_W1215_PCA4_16S_ribosomal_RNA_partial_sequence
GAGCGCC--CCGCAAG-----GGGAGCGGCAGACGGGTGAGTAACCGTGGGG
GATGTACCCGAAGGTACGGAACAACTCCGGAAACTGGAGCTAATACCGTATGTGCCCGA
AAGGGGGAAAGA-----TTTATGCCCTTGGATCAACCCGCGTCAGATTAGC

TAGTTGGTGAGGGTAATGGCCTACCAAGGCAGCATCTGTAGCTGGTCTGAGAGGATGAT
CAGCCACACTGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGAATAT
TGGACAATGGCGCAAGCCTGATCCAGCCATGCCCGTGAGTGATGACGGCCTAGGGTT
GTAAAGCTTTCGACGGGACGA-----TAATGACGGTA
CCCGTAGAAGAACCCCCGGCTAACCTCGTGCAGCAGCCGGTAATACGAAGGGGCTA
GCGTTGTCGGAATCACTGGCGTAAAGCG--CACGTAGGCGATTGTTAAGTCGGGGT
GAAATCCTGAGGCTAACCTCAGAACTGCCCGATACTGGCAATCTGAGTCGGAAAGA
GGTTGGTGGAACAGCTAGTGTAGAGGTGAAATTCTGTAGATATTAGCTAGAACACCAGTGG
CGAAGGCGGCCAAGTGGTCCGGCACTGACGCTGAGGTGCGAAAGCGTGGGAGCAAACAG
GATTAGATAACCTGGTAGTCCACGCCGTAACAGATGGATGCTAGCC-GTTGGGGAGCTTG
CTCTTCGGTGGCGCAGTCAACACATTAAGCATCCCGCTGGGAGTACGGTCGCAAGATT
AAAACCTCAAAGGAATTGACGGGGCCCGACAAGCGGTGGAGCATGTGGTTAATTGAA
GCAACCGCGAGAACCTTACCGCCCTGACATCCCGTGCAGCCTTCAGAGACGAAGGC
CTTCAGTTGGCTGGACCGGAGACAGGTGCTGCATGGCT-GTCGTCAGCTCGTGTGA
GATGTTGGGTTAAGTCCCGAACGAGCGAACCTCGCCCTAGTTGCCATTCAGTT
GGGCACTCTAGGGGACTGCCGGTATAAGCCGAGGAAGGTGGGATGACGTCAAGTC
CTCATGGCCCTTACGGCTGGCTACACACGTGCTACAATGGCGGTGACAGTGGATGCG
AAGGGCGACCCCTAGCAAATCTCAAAAGC-CGTCTCAGTTGCACTCTGCAAC
TCGAGTGCATGAAGGTGGAATCGCTAGTAATCGTAGATCAGCATGCTACGGTAATACGT
TCCCGGCCTGTACACACCGCCGTCACACCATGGAGTTGGTTTACCGAAGGCCT
GCGCCAACC--GCAA

#LC777924.1_Escherichia_coli_VANM2_gene_for_16S_rRNA_partial_sequence
GAACGGTAACAGGAAGCACTGCTTGCTGACGAGTGGCGGACGGGTGAGTAATGTCTGGG
AAACTGCC?GATGGAGGGGATAACTACTGGAAACGGTAGCTAACCGCATAACGTCGC
AAGACCAAAGAGGGGACCTTAGGGCTCTGCCATCGGATGTGCCAGATGGGATTAGC
TAGTAGGTGGGG-TAACGGCTCACCTAGGCGACGATCCCTAGCTGGTCTGAGAGGATGAC
CAGCCACACTGGAACTGAGACACGGTCCAGACTCCTACGGGAGGCAGCAGTGGGAATAT
TGCACAATGGCGCAAGCCTGATGCAGCCATGCCCGTGTATGAAGAAGGCCTCGGGTT
GTAAAGTACTTCAGCGGGAGGAAGGGAGTAAAGTTAACCTTGCTATTGACGTTA
CCCGCAGAAGAACCGGCTAACCTCGTGCAGCAGCCGGTAATACGGAGGGTGC
GCGTTAACCGGAAATTACTGGCGTAAAGCGAACACGCAGGCGTTGTTAAGTCAGATGT
GAAATCCCCGGCTAACCTGGAACTGCATCTGATACTGGCAAGCTTGAGTCTCGTAGA

GGGGGG?AGAATTCCAGGTGTAGCGGTGAAATCGTAGAGATCTGGAGGAATACCGGTGG
CGAAGGC--CC--CTGGACGAAGACTGACGCTCAGGTGCGAAAGCGTGGGAGCAAACAG
GATTAGATAACCCTGGTAGTCACGCCGTAAACGATGTCGACTTGGAGGTTGTGCCCTGA
GGCGTGGCTTCCGGAGCTAACCGTTAACGCGTTAAGTCGACCGCCTGGGAGTACGCCGCAAGGTT
AAAACCTCAAATGAATTGACGGGGGCCGCACAAGCGGTGGAGCATGTGGTTAACCGAT
GCAACCGCGAAGAACCTTACCTGGTCTTGACATCCACGG--AAGTTTCAGAGATGAGAAT
GTGCCTT-CGG-GAACCGTGAGACAGGTGCTGCATGGCTTGTGTCAGCTCGTGTGTGA
AATGTTGGGTTAAGTCCCACGAGCGAACCTTATCCTTGTGCCAGCGTC--CC
GGGAACCTAAAGGAGACTGCCAGTGATAAACTG-GAGGAAGGTGGGATGACGTCAAGTC
ATCATGGCCCTTACGACCAGGGCTACACACGTGCTACAATGGCGCATACAAAGAGAACG
ACCTCGCGAGAGCAAGCG?ACCTCATAAAAGTGCCTCGTAGTCCGGATTGGAGTCTGCAAC
T-GACTCCATGAAGTCGAAATCGCTAGTAATCGTGGATCAGAATGCCACGGTGAATACGT
TCCCCGGGCCTGTACACACCGCCCGTCACACCATGGAGTGGGTTGCAAAAGAAGTAGGT
AGCTTAACCT-TCGG

b) Sử dụng phần mềm MEGA12 để thực hiện so sánh nhiều trình tự gen mã hóa 16S rRNA của chủng nghiên cứu và các chủng gần về mặt phân loại. Các bước thực hiện như sau:

- Mở app **MEGA12**, chọn **Align/Edit/build alignment**, chọn **Create a new alignment** → **DNA**, cửa sổ Alignment Explorer hiện ra với một trình tự trống mặc định là Sequence 1, chọn trình tự này và ấn nút Delete
- Đưa các trình tự cần so sánh vào từ một hoặc nhiều file đã lưu ở mục 1a và 3a: Chọn **Edit/Insert sequences from file** để đưa các file trình tự fasta vào.
- Align các trình tự với nhau bằng cách: Trên cửa sổ **Alignment Explorer**, chọn **Edit>Select All**, khi đó tất cả các trình tự được nhập vào trước đó sẽ được highlight. Tiếp theo, chọn **Alignment/Align by ClustalW**, cửa sổ thông số của ClustalW hiện ra, giữ nguyên các thông số mặc định và ấn OK. Sau khi chương trình thực hiện xong việc align trình tự, bạn kiểm tra lại alignment, cắt bỏ hai đầu của alignment sao cho số nucleotide của các trình tự trong alignment tương đối giống nhau.
- Xuất dữ liệu alignment: Vẫn trên cửa sổ **Alignment Explorer**, chọn **Data/Export alignment/MEGA format**, lưu file alignment bạn vừa tạo ra ở vị trí và tên phù hợp.

Copy cửa sổ kết quả so sánh (Alignment explorer) vào ô trống dưới đây.

```

Pairwise Alignment
Sequence 1: 5292521_C3_27F
Sequence 2: 5292522_C3_1492R
Sequence ends allowed to slide over each other
Alignment score: 8111783

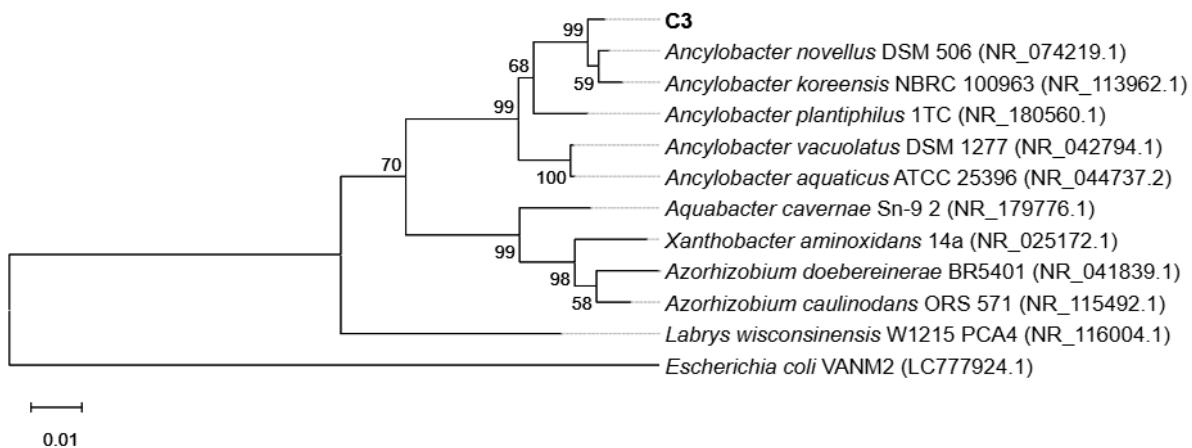
Identities: 0.2747748

```

c) Xây dựng cây phát sinh loài bằng thuật toán Maximum Likelihood hoặc Neighbor-joining với giá trị bootstrap là 1000.

- Tạo cây phát sinh loài: **trên cửa sổ chính** của MEGA12, chọn **Phylogeny/construct/test maximum likelihood tree** hoặc **Phylogeny/construc/test neighbor joining tree**, bạn chọn file chứa các trình tự được align với nhau (MEGA file) để mở file alignment bạn mới tạo ở bước trên. Giữ các thông số Phylogeny reconstruction mặc định, chỉ thay đổi ở **Phylogeny Test: Test of phylogeny** bạn chọn **Standard Bootstrap, Bootstrap replicates**, bạn chọn **1000**.
- Cửa sổ Tree Explorer sẽ xuất hiện sau khi cây phát sinh loài được tạo, bạn thay đổi cây và ấn chuột vào biểu tượng copy to clipboard, paste vào file Powerpoint và tiếp tục chỉnh sửa nếu cần.

Copy cây phát sinh loài vào ô trống dưới đây



Tuần 5: Thiết kế mồi khuếch đại một đoạn gen chỉ thị ở vi sinh vật

4. **Chủng phân lập** được thể hiện một tính trạng nào đó, tính trạng này đã được xác nhận ở một số **chủng vi khuẩn khác** và các gen quy định tính trạng đó đã được xác định trình tự ở các chủng vi khuẩn này. **Bạn có thể thiết lập một giả thuyết rằng chủng phân lập có mang gen có độ tương đồng cao (ortholog) với các gen đã được nghiên cứu ở các chủng khác. Bạn cần chứng minh giả thuyết này bằng cách xác định sự có mặt của ortholog này ở chủng phân lập bằng phản ứng PCR.**

Để làm được điều đó, bạn cần thiết kế mồi cho phản ứng PCR để khuếch đại đoạn gen quan tâm, việc thiết kế mồi dựa vào các gen ở các chủng đã được nghiên cứu, mồi được thiết kế vào vùng bảo thủ (giống nhau) trên các gen ở các chủng đã được nghiên cứu.

Bạn thực hiện thiết kế mồi theo các bước như sau:

a) Tìm kiếm gen quan tâm ở các chủng vi khuẩn đã được nghiên cứu (5-10 ortholog). Chú ý tìm những trình tự mã hóa toàn bộ (hoặc gần như toàn bộ protein quan tâm, và tìm trong những loài gần nhau trong hệ thống phân loại)

Copy các trình tự ortholog của gen quan tâm dưới dạng multi-fasta vào ô trống dưới đây

```
>KX250215.1 Ancylobacter sp. strain TS-1 (aioA) gene, partial cds
GACAACGCCTCGCGTCGACCTGCCAGCAGCAGCAGGGGGCGGACACCCCGGCCTGGTACGCCCGTCCA
TGTACAACATTGTCCGCCAGGACGGCCGAGATGTGCACATCGTCATCAAGCCGACCAGGAATGCTCGGT
GAACCTCCGGCCTCGGATCGGTGCGCGCGCATCGCGAGATGAGCTATTGCGCCAACGCAACACAG
CAGCTCCAGCGCCTAACCGATCCGCTGGCTGGCGCTACGCCAGCTGCAACCGACCAGCTGGACGACG
CGCTCGACCTCGTCGCCGGGTGACATCGCCGTGATGCCGAACAGGGCGAGGATGGCTGTTCGTCTC
CGCTTCGACCATGGCGGCCGGCGGGCTATGAGAACACCTGGGCACCGCAAGCTATTTGGG
GCGATGAAGGTCAAGAACATCCGCATCCACAACCGCCGGCTACAATTCCGAGGTTACGGCTCGCG
ACATGGCGTCGGCGAACTGAACAATTGCTATGAGGACGCCAGCTCGCCGACACCACGTGGCGTCTGG
CACCAACCGCTGGAAACCCAGACCAACTACTTCCTGAACCATTGGTGCGAATTGCGCGCACCTCG
ATGGACAAAAAGAAGGCCGAGTCGGCAGCGAGCCGGTGGCTCAGGGCGGGTGGTCATCGTCGGTCCG
GCCGCACAGTGACGGTCAATGCCTCGGAAGTGGAGGCGGGCAAGGACAACGTGCTTCATCTCGCGCTCAA
TTCCGGCACCGACCTCGCGTTGTTCAACGCTGGTCACCTATGCCGGAAAAGGGCTGGATCGACCGG
GAATTCATCGCGCGTCCACGAAAGACTTCGACAAGGGCGTCGCCAACAAAGGTCTCGATTGCCGAGG
CGCGGAGATCACCGCCCTTCCGAGGCCGACATCGTCAAGGCAGATTACCTGGATGCCGAGCCCCA
```

>KT992350.1 Rhizobium sp. strain YE2-4 (aioA) gene, partial cds

CCACTTCTGCATCGTGGGTGTGGTACACGCCTACACGTGGCGATCAACAAGCAGGGCGGTACCGAT
CCTCAGAAGAACGTGTTGGCGTCGATCTGTCGAACAACAGCAGGCGGAGACTGACGCCCTGGTATTGC
CATCCATGTACAACGTGGTCAAGCAGGATGCCCGACGTTCATATCGTCATCAAGCCGATCATGAATG
CGTGGTCAATTCCGGTCTGGGTCAGTGCGCGGCCAGAATGGCGAAACAAGTTTCCGAAGCGCG
AACACCCAGCAACAGCGTCTGACAGATCCGCTCGTCTGGCGCTACGCCAGATGCAGCCGACGAGCTGG
ACGACGCACTCGATCTGCGCCCGGTGACAGCAAGGATTGTTAAAGACAAGGGTGAGGACGCACTGAT
TGTTCCTGCCTCGATCATGGCGGAGCTGGCGCGGCTATGAAAACACCTGGGCACGGCAAGCTTTAT
TTCGAGGCCATGAAGGTCAAGAATATTCGAATCCACAACCGCCCTGCCTACAATTCCGAGGTCCACGGCA
CGCGCGACATGGCGTCGGCGAATTGAACAATTGCTACGAGGACGCCGACTCGCCGATACGATCGTC
GGTCGGAACCAATGCGCTCGAGACCCAAACCAATTATTCCTAAACCACCTGGTCCCACCTGCGCG
GAAAGCCTCAACAAGAAAAAGCAGGTACTACCGAACGAACCGCATGAGGCTGCCAGAACATCATCGTC
ATCCCGCGCCACTGTGACGATCAACGCCTGTGAGGAAGTGGCCGGCAAGGACAACGTCTGCATTAGC
GATCAACTCCGGCACTGACCTCGCGCTGTTCAATGCACTGTTCACATATATTGCGGACAAGGGCTGGTT
GATCGCGAATTCATCGACAAATCGACCTCGCGCCAAGGCCTGGCAAGGCCGCCAATATCCGCGCG
GCGAATCGGACGCCAACCGGGACATTATCCAACCTCGAGGGGCCCTGAAGGCTGCGTATATCCAT
TGAGGACGCCCGAAAATAACGGGCTCGAGCAAGCCCAGATCGCAAAGCCCGAGTGGATCGCATG
CCGAAGGAAGGCGGAAAGGCCGCCCGTCACTGTTGTTACGAGAACGGCCTCATCTGGGCAACGACA

>KT992346.1 Rhizobium sp. strain K3 (aioA) gene, partial cds

TGTATTGTGGGTGTGGTTATCACGCCTACACGTGGCGATCAACAAGCAGGGCGGTACCGATCCTCAGA
AGAACGTGTTGGCGTCGATCTGTCGAACAACAGCAGGCGGAGACTGACGCCCTGGTATTGCCATCCAT
GTACAACGTGGTCAAGCAGGATGCCCGACGTTCATATTGTCATCAAGCCGATCATGAATCGTGGTC
AATTCCGGCTGGCTCGGTACCGCGCCAGAATGGCGAAACAAGTTTCCGAAGCGCGAACACCC
AGCAACAGCGTCTGACAGATCCGTTGGCTGGCGCTACGCCAGATGCAGCCGACGAGCTGGACGATGC
ACTCGATCTGTCGCCCGGTGACCGCAAAGATCGTAAAGATAAGGGTGAGGACGCACTGATCGTTCT
GCCTCGATCACGGAGGTGCTGGCGCGGCTATGAAAACACCTGGGCACCGCAAGCTTATTCGAGG
CCATGAAGGTCAAGAATATTCGAATGCACAACCGCCCTGCCTACAATTCCGAGGTCCACGGCACGCG
CATGGCGTCGGTGAACTAACAATTGCTACGAGGACGCCGAACCTGCCGATACGATCGTGGCGG
ACCAACCGCGCTCGAGACCCAAACCAACTATTCCTGAACCAACTGGTCCCACCTGCGTGGCGAAAGCC
TCAACAAGAAAAACAGGTACTGCCGAACGAACCGCATGAGGCAGGCCAGAACATCATCGTCGATCCGCG

CCGCACTGTGACGATCAACGCCTGTGAGGAAGTAGCCGGCAAGGACAACGTCTGCATTAGCGATCAAC
TCCGGCACTGACCTCGCTCTGTTCAATGCACTGTTCACATACATCGCGACAAGGGCTGGGTTGATCGCG
AATTCAATCGACAAATCGACCTCTCGTCAAGGCAGGGCGCCGCAATATCCGCACGTGGCGAAC
CGACAGTAAGCCTGGACATTGTCACACTTCGAGGGGCCGTCGAAGGCTGTCGTATATCCATTGAGGAC
GCCCGAAAATAACGGGGCTAGAGCAAGCCGAGATCGTCAAGGCAGGGAGTGGATCGCATGCCAAGG
AAGGCGAAAGCGCCGCCGCGTGAACCTTGCCTGGCCACGGCAATATTGGCCGACCGGGTGGCGCGTCGTT
CACCAACGGCGCCGTGGTGAACCTTGCCTGGCCACGGCAATATTGGCCGACCGGGTGGCGCGTCGTT
CGTTTGGCGGCCACCAAGAAGGATA

>KT992341.1 Rhizobium sp. strain 3AM-13 (aioA) gene, partial cds

CCACTTCTGCATCGTGGATGTGGGTACACGCCTACACGTGGCCGATCAACAAGCAGGGCGGTACCGAT
CCTCAGAAGAATGTATTGGCGTCGATCTGTCGGAACAAACAGCAGGGCGAGACTGACGCCTGGTATTGC
CATCCATGTACAACGTGGTCAAGCAGGATGCCGCGACGTTCATATCGTCATCAAGCCCAGTCATGAATG
CGTGGTCAATTCCGGCCTCGGCTCGTGCGCGGCCAGAATGGCAGAAACGAGTTTCCGAAGCCCG
AACACCCAGCAACAGCGTCTGACAGATCCGGTCTGGCGCTACGGCCAGATGCAGCCAGGACTGGG
ACGATGCACTCGATCTGTCGCCCGGTGACCGCAAAGATCGTAAAGATAAGGGTGAGGACGCACTGAT
CGTTTCTGCCTCGATCACGGAGGTGCTGGCGGGCTATGAAAACACCTGGGCACGGCAAGCTTTAT
TTCGAGGCCATGAAGGTCAAGAATATTCGAATCCACAACCGCCCTGCCTACAATTCCGAGGTCCACGGCA
CGCGCGACATGGCGTCGGTGAATTGAACAATTGCTACGAGGACTCCGAACCGCCGATACGATCGTC
GGTCGGAACCAACCGCCTCGAGACCCAAACCAACTATTCCTGAACCACGGTCCAAACTTGCCTGG
GAAAGCCTCAACAAGAAAAACAGGTACTGCCGAACGAACCGCATGAGGCAGGCCAGAATCATCGTC
ATCCCGCCGCACTGTGACGATCAACGCCTGTGAGGAAGTAGCCGGCAAGGACAACGTCTGCATTAGC
GATCAACTCCGGCACTGACCTCGCTCTGTTCAATGCACTGTTCACATACATCGCGACAAGAGCTGGGTT
GATCGCGAATTCAATCGACAAATCGACCTCTCGTCAGGGCGTGSCACGGCCGCCGCAATATCCGCACGTG
GCGAACCGACAGTAAACCTGGACATTGTCACACTTCGAGGGGCCGTCAGGGCTGTCGTATATCCAT
TGAGGACGCCGCGAAAATAACGGGGCTAGAGCAAGCCGAGATCGTCAAAGCCGCCAGTGGATCGCATG
CCGAAGAAAGGCGGAAAGCGCCGCCGCGTGAATGTCGTTACGAGAAAGGCCATCTGGGCACCGACA

>MF621580.1 Rhizobium sp. strain Cug7 (aioA) gene, partial cds

ATTCCACTTCTGCATCGTGGTTGTGGGTATCACGCCTACACGTGGCCGATCAACAAGCAGGGCGGTAC
GATCCTCAGAAGAACGTGTTGGCGTCGATCTGTCGGAACAAACAGCAGGGCGAGACTGACGCCTGGTATT

CGCCATCCATGTACAACGTGGTCAAGCAGGATGGCCGCACGTTCATATTGTCATCAAGCCGATCATGA
ATGCGTGGTCAATTCCGGCCTGGGCTCGGTACGC GGCCAGAATGGCGAAACAAGTTTCCGAAGCG
CGGAACACCCAGCAACAGCGTCTGACAGATCCGTTGGCTGGCGTACGGCAGATGCAGCGACGAGCT
GGGACGACGCACTCGATCTTGTGCCCGCGTGACCGCAAGGATCGTAAAGATAAGGGTGAGGACGCCCT
GATCGTTCTGCCTTGATCATGGCGGTGCTGGCGGGCTATGAGAACACCTGGGCACCGCAAGCTT
TATTCGAGGCCATGAAGGTCAAGAATATTGAATCCACAACC GCCCTGCCTACAATTGGAGGTGCACG
GCACGCGCACATGGCGTCGGTAGTTGAACAACTGCTACGAAGATGCCAAC TGCCGATACGATCGT
CGCGATCGGTACCAACCGCCTTGAGACCCAAACCAATTATTCCTGAACC ACTGGTCCC AAC TTGCGC
GGCGAAAGCTAAACAAGAAAAAGCAGATACTGCCAACGAAACCGCATGAGGCAGCCAGGATCATCATCG
TTGATCCGCGCCGCACTGTGACGATCAACGCCGTGAGGAAGTAGCCGGCAAGGACAACGTCTGCATT
AGCGATCAACTCCGGCACTGACCTCGCTGTTCAATGCACTGTTCACATACATCGCGAACGGCTGG
GTTGATCGCGAATT CATCGACAAATCGACCTTGC GCAAAGCTTGGCAAGGCCGCGCAATATCCCGCG
GTGGCGAATCGGACGCCAACCGGGACATTATCCAAC TCGAGGGCGCCGTCGAAGGCTGCGTATATC
CATTGAGGACGCCGCGAAAATAACGGGGCTCGAGCAAGCCCAGATCGTCAAAGCCGCCAGTGGATCGC
ATGCCGAAGGAAGGCGAAAGGCCGCCGCTCATGTTGGTTACGAAAAGGAATCATCTGGGCACG
ACAAATCGTCGAC

>MF621579.1 Rhizobium sp. strain Cug6 (aioA) gene, partial cds

ATTCCACTTCTGCATCGTGGGTGTGGTTACAAGGCCTACACCTGGACATCAACCGGCAGGGCGGCACG
AATCCGAGCCAGAACAGTTCAAGGCCGATCTTCCAAGCAGCAGGGTGCCGACAGCAGCGCTGGTACA
CGCCGTCGATGTACAACATCGTCAAGCAGGACGGTAAGGACGTCCACCTGGTCATCATGCCGACAAGAA
CTGCGTGGTGAATT CGGGCCTCGGTT CGGTCCGCCGCGCATGGCGAGACCTCGTACTCAGAGGCG
CGTTGACCCAACAGCAGCGCCTCACGCACCCGATGGCTGGCGTATGGAGCGATGTCCTCGACGTCT
GGGATGATGCTCTCGATCTCGTGGCCCGGGTCACCTGCCAGATCGTCAAGGATCAGGGCGAAGATGGCT
TTTCGTCTCGCTTCGATCATGGCGGCCGGCGGGCTACGAGAACACTTGGGCACCGCAAGCTC
TATTCGGGCCATGAAGGTTAAGAACATCCGGATCCACAACCGCCGGCTACAATTGGAGGTCCACG
CCACGCGCACATGGCATCGCGAGCTAACAACTGCTACGAGGATGCCGAGCTTGGGACACGATCGT
CGTGGTAGGCGCAATCGCTGGAGACCCAGACCAACTATTCCTGAACCATTGGTGCCAACCTGCGC
GGCACGTCGACGGACAAGAAGCGCGCGAGCTCCGAATGAGGCTCATCCACCGCACGGATCGTCA
TCGATCCGCGCGCACAGTCACGGTCAACGCCGTGCGAGGTGGAAGCCGGCAAGGACCGGGTATGCAC
CGCTATCAACTCGGGCAGCGATCTTGCCTGGTCAATGCCTGGATGACGTACATTGCCGAGAAGGTTGG

GTCGACAGGGCGCTGATCGCGGCCTCGACCAACGGGTCGACAAGATGGTCCGGCCAACAAGACGACCC
TGGAGCAGGCTGCCGCTCTGACGGGCCTCACCGTCGACCAGATCCGCCAGTCGGCGGAATGGATCGCGTC
ACCAAAGGAAGGCAATGCGCGACGCCGACCATGTTGCCCTATGAAAAAGGTATCATCTGGGCAACGAC
AAATCGTCGAC

>KT992342.1 *Shinella* sp. strain C23 (aioA) gene, partial cds
CCACTTCTGCATCGTGGGTGCGGGTACCAACGCCTATACTGGCCGATCAACAAGCAGGGCGGTACCGAT
CCGCAGAACGAGTGTGTTGGCGTCGATCTGCGGAACAACAGCAGGCGAAACCGATGCCCTGGTATTGCG
CATCCATGTACAACGTGGTCAAGCAGGATGCCGCGATGTTCATATTGTCATCAAGCCGACCATGAATG
CGTGGTCAATTCCGGCCTCGGCTCGTGCGCGCAAGAACGAGTTGGCTGGCGCTACGGCCAGATGCAGCCACAAGCTGGG
AACACCCAGCAACAGCGTCTGACAGATCCGTTGGCTGGCGCTACGGCCAGATGCAGCCACAAGCTGGG
ACGATGCACTCGATCTGTCGACCGCGTACGCTCGAGGATCGTAAAGACAAGGGTGAGGACGCGCTGAT
TGTTTCTGCCTTCGATCATGGAGGTGCTGGCGCGGCTATGAAAACACCTGGGCACCGCAAGCTTTAT
TTCGAAGCGATGAAGGTCAAGAACATTCGCATCCACAACCGCCCTGCCTACAATTCCGAGGTGCACGGAA
CGCGCGACATGGCGTCGGTGAATTGAACAACGTACGAAGACGCCGAACCTGCCGATACGATCGTGGC
GATCGGGACCAACCGCGTCGAGACGCAAACCAACTATTCCTGAACCATTGGTCCCAAACCTGGGGC
GAAAGCCTTAACAAAAAAAGCAGGTGATGCCAACGAACCGCATGAGGCAGCCAGGATCATTAGTCG
ATCCCGCGCCGCACAGTGACGATCAACGCCCTGCAAGAAGTAGCTGGCAAGGATAATGTTCTGCATCTTGC
GATCAACTCCGGCACTGACCTCGCTCTGTTCAATGCACTGTTCACCTATATGCCGACAAGGGCTGGTC
GATCGGGATTTCATCGACAAATCGACCTTGGCGTCAACCGGTGGCACGGCGCCGCAATATCCCGAGCG
GCGAACATCGGACAGCCAGCCTGGACATTGTCACCTTCGAGGGCGCCGTCGAAGGCTGCGTATGCCAT
CGAGGATGCCGCAAATAACGGGCTCGAGCAAGCCGAGATCGTCAAGGCAGCCGAGTGGATCGGTATG
CCGAAGGAAGGCGGAAAGCGCCGCCGAGTGATGTTGCTACGAAAAGGGCCTCATCTGGGCAACGACA

>MF621576.1 *Shinella* sp. strain Cug3 (aioA) gene, partial cds
ATTCCACTTCTGCATCGTGGGTGCGGGTATCACGCATATACTGGCCGGCAAACAAGCAGGGTGGTACA
GCGCCCGACCGCAATGTTTGGCGTTAACCTGTCAGAACAGCAATATGCGGAATCGGACGCCCTGGTATT
CGCCCTCCATGTACAACATGGTGAAGCAGGACGGTCAGGACGTTCACATTGTTATCAAGCCGGACCAGGC
TTGCGTCGTCAATTCCGGCTGGCTCCGTGCGCGCGCGTATGGCGGAGATGAGTTCTCCGAGGCG
CGCAACACACAGCAGCAGCGCCTGACCGACCCGCTTGTCTGGCGTTACGGCCAGATGCAGCCGACAAGCT
GGGAAGATGCCCTTGACCTTGTGGCGCGCGTTACCGCGGCGTGTCAAAGCGAAGGGCGAGGATGCCCT

CTATGTCTCCGCCTATGATCACGGCGGCCGGCGGCTATGAGAATACTGGGCACGGCAAGCTC
TACTTCGAGGCGATGAAGATCAAGAATATCCGCATCCACAATGCCCGGCATACAACCTCGGAAGTCCATG
CAACGCAGGGACATGGGGTGGCGAGCTAACAAATTGCTACGAGGACGCAGAACTTGCTGACACGATCTT
CTGCGTTGGTGCAAATCGCTGGAAACCCAGACAAACTACTTCCTAACCAACTGGATTCCAATCTCGC
GGGACGACCGCAGCAGAAAAAGACGAGCAGTTGCCGGTGAAGAGCACCCGCTGCCGAATTGTCTTG
TAGACCCGCGTCGTACCGTCAGCATCACGCGCGGAGGCAGAACGCCGGTAAAGACAACGTCTGCATCT
TGCATCAATTCAAGGCACCGACCTTGCTCTGTTAACGCCATTTCAGCCATATCGTCAAAGGGCTGG
GTGGATCAGTCGTTATCGATGCCAGCACGCCGGCTTCACGGAGGCTGTTGCAGCCAACAAGTTGACGC
TCGAGAAGGCCAGTGAGATTGCGCGTCTGGCTTCCGATATCGGAAGGCCACAGTGGATTGCCGA
ACCCAAGGCCAGGCAAGCGCCACGCACCAGCTTCCATGAAAAGGTTCATCTGGGCACAG
AAATCGTCGAC

>AB915361.1 Sinorhizobium sp. KGO-5 (aioA) gene, partial cds
AGGGAGGAACCGAACCCGACCGCAACGCCTCGCCGTCCCACCTCTCGAACGCAGCAAGAGGCCGGAGACCG
ACGCCTGGTACTGCCATCGATGTACAACATCGTAAGCAAGATGGCAGGACGTTCATATCGTCATCAA
GCCGGACCACGAGTGCCTGGTCAATTGGCTTGGCTCCGCGCGCGTATGCCGAGATGAGC
TACTCCGAGGCCCGCAATACACAGCAGCAGCGCCTGACCGATCCCCCTCGTATGGAGATAACGGACAGATGC
AGCCGACGAGTTGGGACGACGCATTGGATCTAGTGGCACGCGTACGGCGACAATAGTCAAGGAAAAGGG
CGAGGATGCCCTCTACGTCTCGGCCTACGACCATGGGGCGCGGGCGGTGGCTACGAGAACACCTGGGGT
ACGGGCAAGCTCTATTGAGGCGATGAAAGTCAAGAATATCCGTATCCACAAACGCCCTGCCTACAAC
CCGAAGTCCACGCCACCGGGACATGGCGTCGGCGAACTTAACAACGTTACGAGGATGCCGAACAG
GGACACGATCTCTGCGTCGGTCTGTAATCCCCCTGAGACACAGACGAACTATTCCCTAACCAACTGGGTC
CCAAACCTGCCGGGATCGACTGCGGACAAGAAGAACGAGCAGCTCCAGGTGAGGAGCATCCGCCAGCAC
GCATCGTGTTCGTCGATCCACGACGCACAGTGACCATCACGCAGCGAACGCCGATGCCGGCTGGAAAA
CGTCCTGCATCTCGCGATCAACTCTGGCACTGACCTTGCTCTGTTAACGCGATCTCACGCACATGTC
GAGAAGGGCTGGATCGACAAGGCCATTGACGCCAGCACAAGCGGA

b) So sánh nhiều trình tự bằng chương trình Clustal Omega của EBI

Copy kết quả so sánh nhiều trình tự vào ô trống dưới đây

CLUSTAL O(1.2.4) multiple sequence alignment

KX250215.1	-----	0
MF621579.1	ATTCACTTCTGCATCGTGGGTGTTACAAGGCCTACACCTGGACATCAACCGGCA	60
KT992342.1	---CCACTTCTGCATCGTGGGTGCGGGTACCACGCCTATACGTGGCGATCAACAAGCA	57
KT992346.1	-----TGTATTGTGGGTGTTATCACGCCTACACGTGGCCGATCAACAAGCA	50
KT992341.1	---CCACTTCTGCATCGTGGGATGTGGTACACGCCTACACGTGGCCGATCAACAAGCA	57
KT992350.1	---CCACTTCTGCATCGTGGGTGTTACACGCCTACACGTGGCCGATCAACAAGCA	57
MF621580.1	ATTCACTTCTGCATCGTGGGTGTTATCACGCCTACACGTGGCCGATCAACAAGCA	60
MF621576.1	ATTCACTTCTGCATCGTGGGTGCGGGTATCACGCATATACTGGCCGGAAACAAGCA	60
AB915361.1	-----A	1

KX250215.1	-----GACAACCGCGTTCGCGCTCGACCTCGCCGAGCAGCAGGGGG	40
MF621579.1	GGGC GG CACGAATCCGAGGCCAGAACAAAG-TTCAAGGCCATTTCCAAGCAGCAGGGTG	119
KT992342.1	GGGC GGT ACCGATCCGAGAACCGTG-TTTGGCGTCGATCTGTCGGAACAAACAGCAGG	116
KT992346.1	GGGC GGT ACCGATCCTCAGAAGAACCGTG-TTTGGCGTCGATCTGTCGGAACAAACAGCAGG	109
KT992341.1	GGGC GGT ACCGATCCTCAGAAGAACCGTG-TTTGGCGTCGATCTGTCGGAACAAACAGCAGG	116
KT992350.1	GGGC GGT ACCGATCCTCAGAAGAACCGTG-TTTGGCGTCGATCTGTCGGAACAAACAGCAGG	116
MF621580.1	GGGC GGT ACCGATCCTCAGAAGAACCGTG-TTTGGCGTCGATCTGTCGGAACAAACAGCAGG	119
MF621576.1	GGGTGGTACAGCGCCCGACCGCAATGTT-TTCGGCGTTAACCTGTCAGAACAGCAATATG	119
AB915361.1	GGGAGGAACCGAACCCGACCGCAACGCCCTCGCCGTCACCTCTCGAACAGCAAGAGG	61

* * * * * * * * *

KX250215.1	CGGACACCCCGGCCTGGTACGCCCGTCCATGTACAACATTGTCCGCCAGGACGGCCGAG	100
MF621579.1	CCGACAGCAGCGCCTGGTACACGCCCGTCAAGCAGGACGGTAAGG	179
KT992342.1	CGGAAACCGATGCCCTGGTATTGCCATCCATGTACAACGTGGTCAAGCAGGATGGCCGCG	176
KT992346.1	CGGAGACTGACGCCCTGGTATTGCCATCCATGTACAACGTGGTCAAGCAGGATGGCCGCG	169
KT992341.1	CGGAGACTGACGCCCTGGTATTGCCATCCATGTACAACGTGGTCAAGCAGGATGGCCGCG	176
KT992350.1	CGGAGACTGACGCCCTGGTATTGCCATCCATGTACAACGTGGTCAAGCAGGATGGCCGCG	176
MF621580.1	CGGAGACTGACGCCCTGGTATTGCCATCCATGTACAACGTGGTCAAGCAGGATGGCCGCG	179
MF621576.1	CGGAATCGGACGCCCTGGTATTGCCCTCATGTACAACATGGTGAAGCAGGACGGTCAGG	179
AB915361.1	CGGAGACCGACGCCCTGGTACTGCCATCGATGTACAACATCGTCAAGCAAGATGGCAGG	121

* * * * * * * * * * * * * *

KX250215.1	ATGTGCACATCGTCATCAAGCCCGACCGCGAATGCTCGGTGAACTCCGGCCTCGGATCGG	160
MF621579.1	ACGTCACCTGGTCATCATGCCGACAAGAACTGCGTGGTGAATTCGGGCCTCGGTTGG	239
KT992342.1	ATGTTCATATTGTCATCAAGCCCGACCATGAATGCGTGGTCAATTCCGGCCTCGGCTCGG	236
KT992346.1	ACGTTCATATTGTCATCAAGCCCGATCATGAATGCGTGGTCAATTCCGGCCTGGGCTCGG	229
KT992341.1	ACGTTCATATCGTCATCAAGCCCGATCATGAATGCGTGGTCAATTCCGGCCTCGGCTCGG	236
KT992350.1	ACGTTCATATCGTCATCAAGCCCGATCATGAATGCGTGGTCAATTCCGGTCTGGGTTCAAG	236
MF621580.1	ACGTTCATATTGTCATCAAGCCCGATCATGAATGCGTGGTCAATTCCGGCCTGGGCTCGG	239
MF621576.1	ACGTTCACATTGTTATCAAGCCGGACCAGGCTTGCCTCGTCAATTCCGGCTCGGCTCCG	239
AB915361.1	ACGTTCATATCGTCATCAAGCCGGACCACGAGTGCCTGGTCAATTCCGGCCTGGCTCCG	181
* *		
KX250215.1	TGCGCGGCGCGCGCATCGCGGAGATGAGCTATTCGCGCCAACGCAACACGCAGCTCCAGC	220
MF621579.1	TCCGCGGCGCCCGCATGGCGGAGACCTCGTACTCAGAGGCGCGTTCGACCCAACAGCAGC	299
KT992342.1	TGCGCGGCGCAAGAATGGCGAACAGAGTTTCCGAAGCGCGAACACCCCAGCAACAGC	296
KT992346.1	TACGCGGGGCCAGAATGGCGAACAGAGTTTCCGAAGCGCGAACACCCCAGCAACAGC	289
KT992341.1	TGCGCGGCCAGAATGGCAGAACAGAGTTTCCGAAGCCGAAACACCCCAGCAACAGC	296
KT992350.1	TGCGCGGCCAGAATGGCGAACAGAGTTTCCGAAGCGCGAACACCCCAGCAACAGC	296
MF621580.1	TACGCGGCCAGAATGGCGAACAGAGTTTCCGAAGCGCGAACACCCCAGCAACAGC	299
MF621576.1	TGCGCGGCCGCGTATGGCGGAGATGAGTTCTCCGAGGCGCGAACACACAGCAGCAGC	299
AB915361.1	TCCGCGGCCGCGTATGGCGGAGATGAGCTACTCCGAGGCCGCAATACACAGCAGCAGC	241
* *		
KX250215.1	GCCTTACCGATCCGCTGGTCTGGCGCTACGCCAGCTGCAACCGACCAGCTGGGACGACG	280
MF621579.1	GCCTCACGCACCCGATGGTCTGGCGCTATGGAGCGATGTCTCCGACGTCTGGGATGATG	359
KT992342.1	GTCTGACAGATCCGTTGGTCTGGCGCTACGCCAGATGCAGCCGACAAGCTGGGACGATG	356
KT992346.1	GTCTGACAGATCCGTTGGTCTGGCGCTACGCCAGATGCAGCCGACGAGCTGGGACGATG	349
KT992341.1	GTCTGACAGATCCGTTGGTCTGGCGCTACGCCAGATGCAGCCGACGAGCTGGGACGATG	356
KT992350.1	GTCTGACAGATCCGCTCGTCTGGCGCTACGCCAGATGCAGCCGACGAGCTGGGACGACG	356
MF621580.1	GTCTGACAGATCCGTTGGTCTGGCGCTACGGCCAGATGCAGCCGACGAGCTGGGACGACG	359
MF621576.1	GCCTGACCGACCCGCTTGTCTGGCGTTACGCCAGATGCAGCCGACAAGCTGGGAAGATG	359
AB915361.1	GCCTGACCGATCCCCCGTATGGAGATAACGGACAGATGCAGCCGACGAGTTGGGACGACG	301
* *		

KX250215.1	TGTTCGTCTCCGCTTCGACCATGGCGGCCGGCGGGCTATGAGAACACCTGGGGCA	400
MF621579.1	TTTCGTCCTCCGCTTCGATCATGGCGGCCGGCGGGCTACGAGAACACTGGGGCA	479
KT992342.1	TGATTGTTCTGCCTCGATCATGGAGGTGCTGGCGGGCTATGAAAACACCTGGGGCA	476
KT992346.1	TGATCGTTCTGCCTCGATCACGGAGGTGCTGGCGGGCTATGAAAACACCTGGGGCA	469
KT992341.1	TGATCGTTCTGCCTCGATCACGGAGGTGCTGGCGGGCTATGAAAACACCTGGGGCA	476
KT992350.1	TGATTGTTCTGCCTCGATCATGGCGGAGCTGGCGGGCTATGAAAACACCTGGGGCA	476
MF621580.1	TGATCGTTCTGCCTTGATCATGGCGGTGCTGGCGGGCTATGAGAACACCTGGGGCA	479
MF621576.1	TCTATGTCTCCGCCTATGATCACGGCGGCCGGCGGGCTATGAGAATACCTGGGGCA	479
AB915361.1	TCTACGTCTCGGCCTACGACCATGGCGGCCGGCGGTGGCTACGAGAACACCTGGGGTA	421

KX250215.1	CCGGCAAGCTCTATTTGGGCGATGAAGGTCAAGAACATCCGCATCCACAACCGCCCCGG	460
MF621579.1	CCGGCAAGCTCTATTCGGGCCATGAAGGTTAAGAACATCCGGATCCACAACCGGCCGG	539
KT992342.1	CCGGCAAGCTTATT CGAAGCGATGAAGGTCAAGAACATCCGCATCCACAACCGCCCTG	536
KT992346.1	CCGGCAAGCTTATT CGAGGCCATGAAGGTCAAGAACATCCGAATGCACAACCGCCCTG	529
KT992341.1	CGGGCAAGCTTATT CGAGGCCATGAAGGTCAAGAACATCCGCATCCACAACCGCCCTG	536
KT992350.1	CGGGCAAGCTTATT CGAGGCCATGAAGGTCAAGAACATCCGCATCCACAACCGCCCTG	536
MF621580.1	CCGGCAAGCTTATT CGAGGCCATGAAGGTCAAGAACATCCGCATCCACAACCGCCCTG	539
MF621576.1	CGGGCAAGCTCTACTCGAGGCCATGAAGGTCAAGAACATCCGCATCCACAATCGCCCGG	539
AB915361.1	CGGGCAAGCTCTATT CGAGGCCATGAAGGTCAAGAACATCCGTATCCACAACCGCCCTG	481
*	*****	*

KX250215.1	CCTACAATTCCGAGGTTCACGGCTCGCGCAGACATGGGCGTCGGCGAACTGAACAAATTGCT	520
MF621579.1	CCTACAATTCCGAGGTCACGCCACGCGCAGACATGGGCATCGCGAGCTAACAACTGCT	599
KT992342.1	CCTACAATTCCGAGGTGACGGAACCGCGCAGACATGGGCGTCGGTGAATTGAACAACTGCT	596
KT992346.1	CCTACAATTCCGAGGTCCACGGCACGCGCAGACATGGGCGTCGGTGAACTGAACAAATTGCT	589
KT992341.1	CCTACAATTCCGAGGTCCACGGCACGCGCAGACATGGGCGTCGGTGAATTGAACAAATTGCT	596
KT992350.1	CCTACAATTCCGAGGTCCACGGCACGCGCAGACATGGGCGTCGGCGAATTGAACAAATTGCT	596
MF621580.1	CCTACAATTCCGAGGTGACGGCACGCGCAGACATGGGCGTCGGTGAGTTGAACAACTGCT	599
MF621576.1	CATACAACCTCGGAAGTCCATGCAACGCGGGACATGGGGGTGGCGAGCTAACAAATTGCT	599
AB915361.1	CCTACAACCTCGGAAGTCCACGCCACCCGGGACATGGGCGTCGGCGAACTTAACAACTGTT	541
* * * * *		

KX250215.1	ATGAGGACGCCAGCTGCCGACACCATCGGCCGTGGCACCAACCGCTGGAAACCC	580
MF621579.1	ACGAGGATGCCGAGCTTGC GGACACGATCGTGGTAGGC GCAATCCGCTGGAGACCC	659
KT992342.1	ACGAAGACGCCGAACTTGCCGATACGATCGTGGC ATCGGGACCAACCGCCTCGAGACGC	656
KT992346.1	ACGAGGACGCCGAACTTGCCGATACGATCGCGGT CGGAACCAACCGCCTCGAGACCC	649
KT992341.1	ACGAGGACTCCGAACTTGCCGATACGATCGCGGT CGGAACCAACCGCCTCGAGACCC	656
KT992350.1	ACGAGGACGCCGAACTCGCCGATACGATCGCGGT CGGAACCAATCGCCTCGAGACCC	656
MF621580.1	ACGAAGATGCCGAACTTGCCGATACGATCGCGATCGGTACCAACCGCTTGAGACCC	659
MF621576.1	ACGAGGACGCAGAACTTGCTGACACGATCTCTGC GTTGGTGCAAATCCGCTGGAAACCC	659
AB915361.1	ACGAGGATGCCGAACTAGCGGACACGATCTCTGC GTCGGTGCTAACCGCTTGAGACAC	601

KX250215.1	AGACCAACTACTCCTGAACCATTGGGTGCCGATTCGCGCGCACCTCGATGGACAAAA	640
MF621579.1	AGACCAACTATTCCTGAACCATTGGGTGCCGACCTGCGCGCACGTCGACGGACAAGA	719
KT992342.1	AAACCAACTATTCCTGAACCATTGGTCCCAAACCTGCGGGCGAAAGCCTAACAAAA	716
KT992346.1	AAACCAACTATTCCTGAACCACTGGTCCCAAACCTGCGTGGCGAAAGCCTAACAGA	709
KT992341.1	AAACCAACTATTCCTGAACCACTGGTCCCAAACCTGCGTGGCGAAAGCCTAACAGA	716
KT992350.1	AAACCAATTATTCCTAAACCACTGGTCCCAAACCTGCGCGCGAAAGCCTAACAGA	716
MF621580.1	AAACCAATTATTCCTGAACCACTGGTCCCGAACCTGCGCGCGAAAGCCTAACAGA	719
MF621576.1	AGACAAACTACTCCTCAACCACTGGATTCCAATCTCGCGGGACGACCGCAGCGAAAA	719
AB915361.1	AGACGAACTATTCCTCAACCACTGGTCCCAAACCTGCGGGGATCGACTGCAGACAAGA	661

KX250215.1	AGAAGGCCGAGTTCGGCAGCGAGCCGGTGGCTCAGGGCCGGTGGTCATCGTCGGTCCGC	700
MF621579.1	AGCGCGCCGAGCTTCCGAATGAGGCTCATCCACCGCACGGATCGTCATCATCGATCCGC	779
KT992342.1	AAAAGCAGGTGATGCCAACGAACCGCATGAGGCAGGCCAGGATCATTATAGTCGATCCGC	776
KT992346.1	AAAAACAGGTACTGCCGAACGAACCGCATGAGGCAGCCAGAATCATCATCGTCGATCCGC	769
KT992341.1	AAAAACAGGTACTGCCGAACGAACCGCATGAGGCAGCCAGAATCATCATCGTCGATCCGC	776
KT992350.1	AAAAGCAGGTACTACCGAACGAACCGCATGAGGCAGCCAGGATCATCATCGTCGATCCGC	776
MF621580.1	AAAAGCAGATACTGCCAACGAACCGCATGAGGCAGCCAGGATCATCATCGTTGATCCGC	779
MF621576.1	AAGACGAGCAGTTGCCGGGTAAAGAGCACCCGCCTGCCGAATTGTCTTGAGACCCGC	779
AB915361.1	AGAACGAGCAGCTTCCAGGTGAGGAGCATCCGCCAGCACGCATCGTGTGTCGATCCAC	721
	* * * * * * * * * * * * *	

KX250215.1	GCCGCACAGTGACGGTCAATGCCTGCGAAGTGGAGGCAGGACAACGTGCTTCATC	760
MF621579.1	GGCGCACAGTCACGGTCAACGCCTGCGAGGTGGAAGCCGGCAAGGACCGGGTGTGACCC	839
KT992342.1	GCCGCACAGTGACGATCAACGCCTGCGAAGAAGTAGCTGGCAAGGATAATGTTCTGCATC	836
KT992346.1	GCCGCACTGTGACGATCAACGCCTGTGAGGAAGTAGCCGGCAAGGACAACGTCCGCATT	829
KT992341.1	GCCGCACTGTGACGATCAACGCCTGTGAGGAAGTAGCCGGCAAGGACAACGTCCGCATT	836
KT992350.1	GCCGCACTGTGACGATCAACGCCTGTGAGGAAGTAGGGCCGGCAAGGACAACGTCCGCATT	836
MF621580.1	GCCGCACTGTGACGATCAACGCCTGTGAGGAAGTAGCCGGCAAGGACAACGTCCGCATT	839
MF621576.1	GTCGTACCGTCAGCATCAACGCCGGAGGCAGAACCGGTAAGACAACGTCCGCATT	839
AB915361.1	GACGCACAGTGACCATCAACGCAGCCGAAGCCGATGCCGCTGGAAAACGTCCGCATT	781
	* * * * * * * * * * * * *	

KX250215.1	TGCGGCTCAATTCCGGCACCGACCTCGCGTTGTTCAACGCCTGGTTCACCTATGCGCGG	820
MF621579.1	TCGCTATCAACTCGGGCAGCGATCTGCGCTGTTCAATGCCTGGATGACGTACATTGCCG	899
KT992342.1	TTGCGATCAACTCCGGCACTGACCTCGCTCTGTTCAATGCACTGTTCACCTATATGCCG	896
KT992346.1	TAGCGATCAACTCCGGCACTGACCTCGCTCTGTTCAATGCACTGTTCACATACATCGCG	889
KT992341.1	TAGCGATCAACTCCGGCACTGACCTCGCTCTGTTCAATGCACTGTTCACATACATCGCG	896
KT992350.1	TAGCGATCAACTCCGGCACTGACCTCGCGCTGTTCAATGCACTGTTCACATATATTGCGG	896
MF621580.1	TAGCGATCAACTCCGGCACTGACCTCGCTCTGTTCAATGCACTGTTCACATACATCGCG	899
MF621576.1	TTGCGATCAATTAGGCACCGACCTTGCTCTGTTCAACGCCATTTCAGCCATATCGTCG	899
AB915361.1	TCGCGATCAACTCTGGCACTGACCTTGCTCTGTTCAACGCAGCCACATCGTCG	841
	* * * * * * * * * * * * *	

KX250215.1	AAAAGGGCTGGATCGACCGGAATTCATCGGCGGTCCACGAAAGACTTCGACAAGGC GG	880
MF621579.1	AGAAGGGTTGGGTCGACAGGGCCTGATCGCGGCCGACCAACGGGTCGACAAGATGG	959
KT992342.1	ACAAGGGCTGGGTCGATCGGAATTCATCGACAAATCGACCTTGC GTCAACCGTG GAC	956
KT992346.1	ACAAGGGCTGGGTTGATCGGAATTCATCGACAAATCGACCTCTCGTCAAGGC GTGGCAC	949
KT992341.1	ACAAGAGCTGGGTTGATCGGAATTCATCGACAAATCGACCTCTCGTCAAGGC GTGSCAC	956
KT992350.1	ACAAGGGCTGGGTTGATCGGAATTCATCGACAAATCGACCTCGGCCAAGGC GTGGCAA	956
MF621580.1	ACAAGGGCTGGGTTGATCGGAATTCATCGACAAATCGACCTTGC GCCAAGGCTTGGCAA	959
MF621576.1	AAAAGGGCTGGGTGGATCAGCTTATCGATGCCAGCACGCCGGCTTCACGGAGGCTG	959
AB915361.1	AGAAGGGCTGGATCGACAAGGC GTTCATCGACGCCAGCACAAGCGGA-----	888

* * * * * * * * * * * *

KX250215.1	TCGCCGCCAACAAAGGTCTCGATTGCCGAGGC GGGGAGATCACCGCCTTCCGAG---	936
MF621579.1	TCGCGGCCAACAAAGACGACCCCTGGAGCAGGCTG-----	992
KT992342.1	GGCCGCCGCAATATCCCGAGCGGGAATCGGACAGCCAGCCTGGACATTGTCCA ACT	1016
KT992346.1	GGCCGCCGCAATATCCGCACGTGGGAATCCGACAGTAAGCCTGGACATTGTCCA ACT	1009
KT992341.1	GGCCGCCGCAATATCCGCACGTGGGAATCCGACAGTAACCTGGACATTGTCCA ACT	1016
KT992350.1	GGCCGCCGCAATATCCGCACGTGGGAATCGGACGCCAACCGGGACATTATCCA ACT	1016
MF621580.1	GGCCGCCGCAATATCCGCACGTGGGAATCGGACGCCAACCGGGACATTATCCA ACT	1019
MF621576.1	TTGCAGCCAACAAGTTGACGCTCG-AG--AAGGCCAGTGAGATTGCG-----	1004
AB915361.1	-----	888

KX250215.1	-----GCCGACATCGTCAAGGC GATTACCTGGATGCCGAGCCCCA-----	977
MF621579.1	-----CCGCTCTGACGGGCC	1007
KT992342.1	TCGAGGGCGCCGTCGAAGGCTGTCGTATGTCCATCGAGGATGCCG GAAAATAACGGGGC	1076
KT992346.1	TCGAGGGCGCCGTCGAAGGCTGTCGTATATCCATTGAGGACGCC GCGAAAATAACGGGGC	1069
KT992341.1	TCGAGGGCGCCGTCGAAGGCTGTCGTATATCCATTGAGGACGCC GCGAAAATAACGGGGC	1076
KT992350.1	TCGAGGGCGCCGTCGAAGGCTGTCGTATATCCATTGAGGACGCC GCGAAAATAACGGGGC	1076
MF621580.1	TCGAGGGCGCCGTCGAAGGCTGTCGTATATCCATTGAGGACGCC GCGAAAATAACGGGGC	1079
MF621576.1	-----GCG	1007
AB915361.1	-----	888

KX250215.1	-----	977
MF621579.1	TCACCGTCGACCAGATCCGCCAGTCGGCGGAATGGATCGCGTACCAAAGGAAGGCAATG	1067
KT992342.1	TCGAGCAAGCCGAGATCGTCAAGGCAGCCGAGTGGATCGGTATGCCGAAGGAAGGCGAA	1136
KT992346.1	TAGAGCAAGCCGAGATCGTCAAGGCAGCCGAGTGGATCGCATGCCGAAGGAAGGCGAA	1129
KT992341.1	TAGAGCAAGCCGAGATCGTCAAAGCCGCCAGTGGATCGCATGCCGAAGAAAGGCGAA	1136
KT992350.1	TCGAGCAAGCCCAGATCGTCAAAGCCGCCAGTGGATCGCATGCCGAAGGAAGGCGAA	1136
MF621580.1	TCGAGCAAGCCCAGATCGTCAAAGCCGCCAGTGGATCGCATGCCGAAGGAAGGCGAA	1139
MF621576.1	TCTCGGTTCCGATATCGCAAGGCGGCACAGTGGATTGCCAACCCAAGGCCGACGGCA	1067
AB915361.1	-----	888

KX250215.1	-----	977
MF621579.1	CGCGACGCCGCACCATGTTCGCCTATGAAAAGGTATCATCTGGGCAACGACAAATCGT	1127
KT992342.1	AGCGCCGCCGAGTGTATGTTCGGTTACGAAAAGGGCCTCATCTGGGCAACGACA-----	1190
KT992346.1	AGCGCCGCCGCGTGATGTTCGGTTACGAAAAGGCCTGATCTGGGCAACGATAACTACC	1189
KT992341.1	AGCGCCGCCGCGTCATGTTGGTTACGAGAAAGGCCTCATCTGGGCAACGACA-----	1190
KT992350.1	AGCGCCGCCGCGTCATGTTGGTTACGAGAAGGGCCTCATCTGGGCAACGACA-----	1190
MF621580.1	AGCGCCGCCGCGTCATGTTGGTTACGAAAAGGAATCATCTGGGCAACGACAAATCGT	1199
MF621576.1	AGCGCCGACGCACCATGCTCCCTATGAAAAGGTTCATCTGGGCAACGACAAATCGT	1127
AB915361.1	-----	888

KX250215.1	-----	977
MF621579.1	CGAC-----	1131
KT992342.1	-----	1190
KT992346.1	GCACCAACGGCGCCGTGGTGAACCTGCGCTGCCACGGCAATTGGCCGACCGGGTG	1249
KT992341.1	-----	1190
KT992350.1	-----	1190
MF621580.1	CGAC-----	1203
MF621576.1	CGAC-----	1131
AB915361.1	-----	888

KX250215.1	-----	977
MF621579.1	-----	1131
KT992342.1	-----	1190
KT992346.1	GCGCGTCGTTCGTTGGCGGCCACCAAGAAGGATA	1286
KT992341.1	-----	1190
KT992350.1	-----	1190
MF621580.1	-----	1203
MF621576.1	-----	1131
AB915361.1	-----	888

c) Xác định các vị trí bảo thủ trên đoạn gen, ghi lại vị trí của các vùng bảo thủ đó với mỗi gen đưa vào so sánh

Vùng gen bảo thủ ở vị trí nào

Phân tích đa trình tự cho thấy hai vùng bảo tồn nổi bật nằm tại vị trí nucleotide 103–122 và 403–422 trên trình tự gen aioA của Ancylobacter sp. TS-1, tương ứng với các đoạn trình tự GTGCACATCGTCATCAAGCC và GGCAAGCTCTATTTGGGGC.

5. Thiết kế mồi bằng Primer-Blast (NCBI)

Sử dụng trình tự của toàn bộ gen quan tâm ở một chủng gần gũi nhất về mặt phân loại với chủng nghiên cứu, đưa vào làm khuôn cho việc thiết kế mồi trên Primer-Blast.

Lựa chọn các cặp mồi phù hợp và copy kết quả vào ô trống dưới đây

Primer pair 1								
	Sequence (5'→3')	Template strand	Length	Start	Stop	Tm	GC%	Self complementarity
Forward primer	GTGCACATCGTCATCAA	Plus	17	103	119	52.32	47.06	6.00
Reverse primer	GCCCCAAATAGAGCTTG	Minus	18	422	405	53.34	50.00	4.00
Product length	320							2.00

6. Phân tích các đặc tính của protein mà gen đó mã hóa, từ đó dự đoán được một phần tính chất của protein trên chủng nghiên cứu. Các đặc tính này bao gồm:

a) Trình tự protein mà gen mã hóa: sử dụng công cụ Translate trên EXPASY

Kết quả:

DNAFGVDLAEQQGADTPAWYAPSMYNIVRQDGRDVHIVIKPDRECSVNSGLGSVRGARIAEMSYS
RQRNTQLQLRTDPLVWRYGQLQPTSWDDALDLVARVTSAVIAEQGEDGLFVSADFHDGGAGGGYE

NTWGTGKLYFGAMKVKNIRIHNRPAYNSEVHGSRDMGVGELNNCYEDAQLADTIVAVGTNALETQ
TNYFLNHWPVNLRGTSMDKKAEFGSEPVAQGRVVIVGPRRTVTVNACEVEAGKDNVLHLALNS
GTDLALFNAWFTYAAEKGWIDREFIGASTKDFDKAVAANKVSIAEAAEITGLSEADIVKAITWIAEP

Results of translation

- Open reading frames are highlighted in red
- Select your initiator on one of the following frames to retrieve your amino acid sequence

[Download all the translated frames](#)

5'3' Frame 1

DNAFGVDLAEQQGADTPAWYAPS**M**YNIVRQDGRDVHIVIKPDRECSVNSGLGSVRGARIAE**M**SYSRQRNTQLQRLTDPLWRYGQLQPTSWDDALDLVARTSAVIAEQGEDGLFVSAFDHGGAGGGYENTWGTGKLYFGA**M**KVKNIRIHNRPAYNSEVHGSRDMGVGELNNCYEDAQLADTIVAVGTNALETQTNYFLNHWPVNLRGTSMDKKAEFGSEPVAQGRVVIVGPRRTVTVNACEVEAGKDNVLHLALNSGTDLALFNAWFTYAAEKGWIDREFIGASTKDFDKAVAANKVSIAEAAEITGLSEADIVKAITWIAEP

b) Xác định khối lượng phân tử (molecular weight - Mw), điểm đắng điện (isoelectric point - PI) của protein đó.

Nếu protein sau khi dịch mã có đoạn tín hiệu vận chuyển qua màng, đoạn trình tự này sẽ bị cắt bỏ sau khi được vận chuyển qua màng tế bào, khi xác định khối lượng phân tử và điểm đắng điện của protein nêu trên, cần sử dụng trình tự protein sau khi đã cắt bỏ trình tự tín hiệu. Tính toán một cách tương đối, khối lượng phân tử của protein có thể được xác định như sau: một amino acid có khối lượng khoảng 110 Da, nếu một protein có 100 aminoacid, nó có khối lượng phân tử xấp xỉ là 11 kDa.

Kết quả:

Theoretical pl/Mw: 5.08 / 35444.57 (Expansy)

Compute pl/Mw - Results

Theoretical pl/Mw (average)

Sequence:

10	20	30	40	50	60
DNAFGVDLAE	QQGADTPAWY	APSMYNIVRQ	DGRDVHIVIK	PDRECSVNSG	LGSVRGARIA
70	80	90	100	110	120
EMSYSRQRNT	QLQRLTDPLV	WRYGQLQPTS	WDDALDLVAR	VTSAVIAEQG	EDGLFVSAFD
130	140	150	160	170	180
HGGAGGGYEN	TWGTGKLYFG	AMKVKNIRIH	NRPAYNSEVH	GSRDMGVGEL	NNCYEDAQLA
190	200	210	220	230	240
DTIVAVGTNA	LETQTNYFLN	HWVNLRGTS	MDKKAEFGS	EPVAQGRVVI	VGPRRTVTVN
250	260	270	280	290	300
ACEVEAGKDN	VLHLALNSGT	DLALFNAWFT	YAAEKGWIDR	EFIGASTKDF	DKAVAANKVS
310	320				
IAEAAEITGL	SEADIVKAIT	WIAEP			

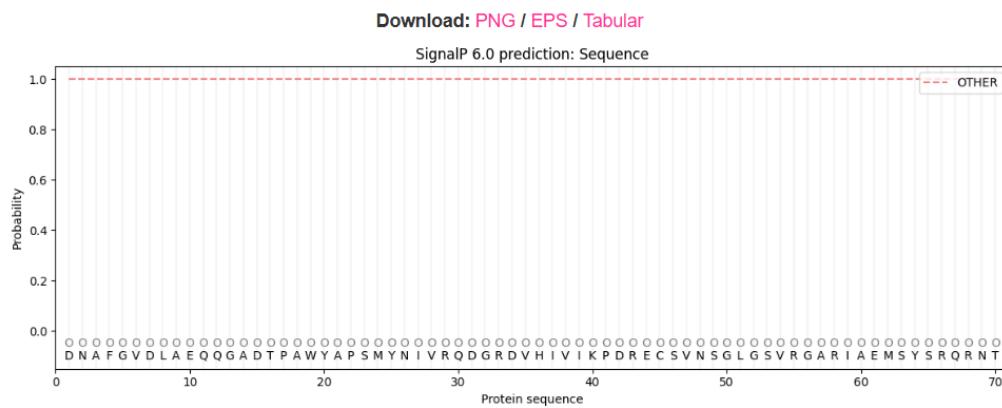
Theoretical pl/Mw: 5.08 / 35444.57

c) Quá trình cải biến sau dịch mã (post-translation modification) xảy ra ở một số loại protein và chuyển hóa tiền-protein thành dạng có hoạt tính hoặc vận chuyển protein đến vị trí mà tại đó nó có hoạt tính xúc tác (ví dụ như trên màng tế bào, hoặc trong periplasmic). Một trong các quá trình cải biến protein sau dịch mã là vận chuyển protein qua màng tế bào, thông thường, các protein này có một đoạn trình tự peptide làm tín hiệu (signal peptide), sau khi protein được vận chuyển qua màng ra môi trường hoặc ra periplasmic, đoạn trình tự này bị cắt bỏ. Sử dụng công cụ SignalP 6.0

(<https://services.healthtech.dtu.dk/services/SignalP-6.0/>) để xác định đoạn trình tự protein sau quá trình dịch mã có đoạn trình tự tín hiệu cho việc vận chuyển qua màng tế bào không

Kết quả: Không

Sequence Prediction: Other		Protein type	Signal Peptide (Sec/SPI)	Lipoprotein signal peptide (Sec/SPII)	TAT signal peptide (Tat/SPI)	TAT Lipoprotein signal peptide (Tat/SPII)	Pilin-like signal peptide (Sec/SPIII)
Likelihood	1	0	0	0	0	0	0



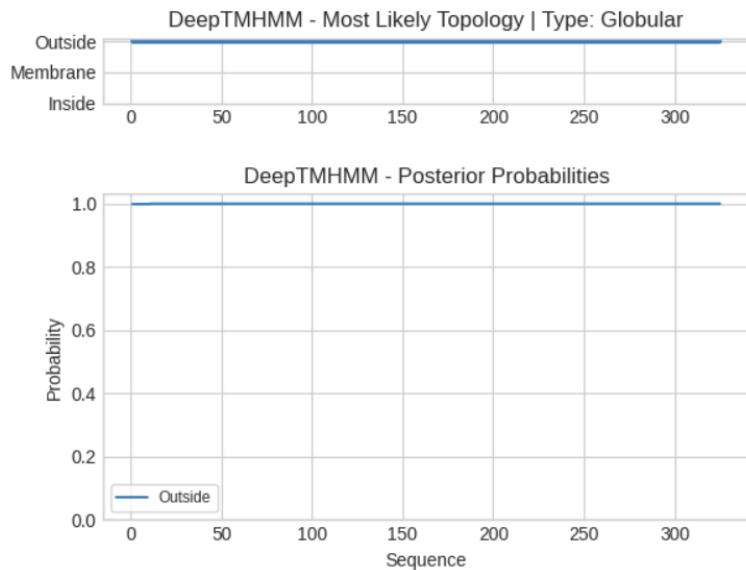
d) Tìm công cụ online DeepTMHMM-1.0

(<https://services.healthtech.dtu.dk/services/DeepTMHMM-1.0/>) để xác định xem protein này có gắn trên màng tế bào (transmembrane) không?

Kết quả: Không

DeepTMHMM - Predictions

Predicted topologies can be downloaded in [.gff3 format](#) and [.3line format](#)



- e) Protein được mã hóa bởi gen đó đã được xác định cấu trúc bằng thực nghiệm chưa? Nếu có, bạn hãy tìm cấu trúc trên protein databank hoặc NCBI và tìm hiểu về nó. Nếu chưa, bạn sử dụng AlphaFold server để dự đoán cấu trúc của nó.

Kết quả:

